

Metoda największej wiarygodności i informacja Fisher'a w fizyce i ekonofizyce

(Maximum likelihood method and Fisher's information in physics and econophysics)

Jacek Syska¹

Institute of Physics, University of Silesia, Uniwersytecka 4, 40-007 Katowice, Poland

Skrypt dla studentów ekonofizyki²

¹jacek.syska@us.edu.pl

²Współfinansowanie projektu "Uniwersytet partnerem gospodarki opartej na wiedzy" w ramach Europejskiego Funduszu Społecznego. Kapitał Ludzki Narodowa Strategia Spójności.

Spis treści

Abstract	4
Wstęp	5
1 Metoda największej wiarygodności	9
1.1 Podstawowe pojęcia MNW	9
1.2 Wnioskowanie w MNW	13
1.2.1 Wiarygodnościowy przedział ufności	13
1.2.2 Rozkłady regularne	16
1.2.3 Weryfikacja hipotez z wykorzystaniem ilorazu wiarygodności	16
1.3 MNW w analizie regresji	18
1.3.1 Dewiancja jako miara dobroci dopasowania. Rozkład Poissona.	20
1.3.2 Analiza regresji Poissona.	22
1.3.2.1 Test statystyczny dla doboru modelu w regresji Poissona	24
1.3.2.2 Podobieństwo dewiancji do SKR analizy częstotliwościowej	26
1.4 Zasada niezmienniczości ilorazu funkcji wiarygodności	28
2 Entropia względna i informacja Fishera	29
2.1 Obserwowana i oczekiwana informacja Fishera	29
2.1.1 Wartość oczekiwana i wariancja funkcji wynikowej	33
2.2 Wstęp do geometrii różniczkowej na przestrzeni statystycznej i α -koneksja	35
2.2.1 Przestrzeń statystyczna dualnie płaska	42
2.2.2 Dualne układy współrzędnych	44
2.2.2.1 Transformacja Legendre'a pomiędzy parametryzacjami dualnymi	46
2.2.3 Geometryczne sformułowanie teorii estymacji dla EFI	49
2.2.4 Uwaga o rozwinięciu rozkładu w szereg Taylora	49
2.3 Twierdzenie Rao-Cramera i DORC	52
2.3.1 Skalarne Twierdzenie Rao-Cramera	52
2.3.1.1 Dowód TRC (wersja dowodu dla przypadku skalarnego)	53
2.3.1.2 Przykład skalarny DORC dla rozkładu normalnego	53
2.3.2 Wieloparametrowe Twierdzenie Rao-Cramera	54
2.3.2.1 Pierwsza wersja wieloparametrowego TRC	55
2.3.2.2 Przykład wektorowego DORC	55
2.3.2.3 Druga wersja wieloparametrowego TRC	57
2.4 Entropia informacyjna Shannona i entropia względna	58
2.4.1 Interpretacja entropii Shannona	59

2.4.2	Przypadek ciągłego rozkładu prawdopodobieństwa	59
2.4.3	Entropia względna jako miara odległości	61
2.5	Geometria przestrzeni rozkładów prawdopodobieństwa i metryka Rao-Fishera	63
2.6	Informacja Fishera	67
2.6.1	Informacja Fishera jako entropia	67
2.7	Pojęcie kanału informacyjnego	71
2.7.1	Pojemność informacyjna dla zmiennej losowej położenia	72
2.8	Pomiar układu w podejściu Friedena-Soffera	77
2.8.1	Przykład: Estymacja w fizycznych modelach eksponentialnych	79
3	Zasady informacyjne	84
3.1	Estymacja w statystyce klasycznej a estymacja fizyczna. Postawienie problemu	84
3.1.1	Strukturalna zasada informacyjna. Metoda EFI	85
3.1.1.1	Całka rozwinięcia Taylora	87
3.1.1.2	I oraz Q dla parami niezależnych zmiennych położeniowych próby	88
3.2	Przepływ informacji	89
3.3	Kinematyczna postać informacji Fishera	93
3.3.1	Postać kinematyczna pojemności zapisana w prawdopodobieństwie	97
3.4	Równania master	97
3.5	Podsumowanie rozważań	98
4	Kryteria informacyjne w teorii pola	100
4.1	Informacja Fishera i klasyfikacja modeli	100
4.1.1	Podział modeli ze względu na N oraz kategorie I	101
4.1.1.1	Dowód podziału na dwie kategorie I	101
4.1.2	Podział modeli ze skończonym I na podklasy z różnym N	102
4.1.3	Konkluzje i konsekwencje podziału modeli na kategorie I	102
4.2	Równania różniczkowe metody EFI	103
4.2.1	Ogólna postać funkcji gęstości TFI oraz obserwowane zasady informacyjne	103
4.2.1.1	Postać obserwowana zasad informacyjnych	105
4.3	TFI oraz specyficzne formy Q w teorii pola	106
4.3.1	Informacja Fouriera	107
4.3.1.1	Informacja Fouriera dla amplitudy rzeczywistej	108
4.3.2	Skalary Kleina-Gordona	110
4.3.3	TFI równania Kleina-Gordona dla pól rangi N	111
4.3.3.1	Wstępna foliacja \mathcal{S} oraz pochodna kowariantna. Ogólny zarys problemu	111
4.3.3.2	Równanie ruchu Diraca dla pola swobodnego rangi N .	113
4.3.4	Końcowe uwagi o wkładzie Q w zasadę strukturalną	117
4.3.5	Zasada najmniejszego działania v.s. zasady informacyjne	117
5	Przykłady z fizyki statystycznej i ekonofizyki oraz efekt EPR-Bohm'a	120
5.1	Wyprowadzenie klasycznej fizyki statystycznej z informacji Fishera	120
5.1.1	Fizyczne sformułowanie zagadnienia	121
5.1.2	Informacja kinetyczna i strukturalna oraz sformułowanie zasad informacyjnych	121
5.1.3	Rozkład Boltzmanna dla energii	123
5.1.4	Model Aoki-Yoshikawy dla produktywności branż	131

5.1.5	Rozkład Maxwella-Boltzmann dla prędkości	136
5.1.6	Informacja Fishera jako ograniczenie dla wzrostu entropii	141
5.1.6.1	Wynik dla strumienia cząstek ze spinem $1/2$	144
5.2	Zastosowanie wprowadzonego formalizmu do analizy paradoksu EPR	145
5.2.1	Warunki brzegowe	146
5.2.2	Pojemność informacyjna dla zagadnienia EPR-Bohm'a	149
5.2.3	Informacja strukturalna. Amplituda prawdopodobieństwa	152
5.2.4	Niepewność wyznaczenia kąta	161
5.2.4.1	Wpływ zaszumienia pomiaru	162
5.2.5	Informacja Q jako miara splątania	162
6	Zakończenie	164
7	Dodatki	166
7.1	Dodatek: Zasada nieoznaczoności Heisenberga	166
7.2	Dodatek: Równanie Schrödingera	169
7.3	Dodatek: Rezultaty EFI dla elektrodynamiki Maxwella oraz teorii grawitacji	171
7.3.1	Dodatek: Pole cechowania Maxwella	171
7.3.2	Dodatek: Metoda EFI dla teorii grawitacji	173
7.4	Dodatek: Informacyjny odpowiednik drugiej zasady termodynamiki: Twierdzenie I . . .	175
7.4.1	Dodatek: Temperatura Fishera	176

Abstract

Three steps in the development of the maximum likelihood (ML) method are presented. At first, the application of the ML method and Fisher information notion in the model selection analysis is described (Chapter 1). The fundamentals of differential geometry in the construction of the statistical space are introduced, illustrated also by examples of the estimation of the exponential models.

At second, the notions of the relative entropy and the information channel capacity are introduced (Chapter 2). The observed and expected structural information principle (IP) and the variational IP of the modified extremal physical information (EPI) method of Frieden and Soffer are presented and discussed (Chapter 3). The derivation of the structural IP based on the analyticity of the logarithm of the likelihood function and on the metricity of the statistical space of the system is given.

At third, the use of the EPI method is developed (Chapters 4-5). The information channel capacity is used for the field theory models classification. Next, the modified Frieden and Soffer EPI method, which is a nonparametric estimation that enables the statistical selection of the equation of motions of various field theory models (Chapter 4) or the distribution generating equations of statistical physics models (Chapter 5) is discussed. The connection between entanglement of the momentum degrees of freedom and the mass of a particle is analyzed. The connection between the Rao-Cramér inequality, the causality property of the processes in the Minkowski space-time and the nonexistence of tachions is shown. The generalization of the Aoki-Yoshikawa sectoral productivity econophysical model is also presented (Chapter 5). Finally, the Frieden EPI method of the analysis of the EPR-Bhom experiment is presented. It differs from the Frieden approach by the use of the information geometry methods. This work is destined mainly for students in physics and econophysics. (At present only Polish version is available).

Wstęp

Tematem skryptu jest metoda największej wiarygodności (MNW) oraz informacja Fishera (IF) w fizyce i statystyce. Problem dotyczy bardzo aktualnego sposobu konstrukcji modeli fizycznych, który wywodzi się ze statystycznego opisu zjawisk, którego formalizm pozwala na opis całego spektrum różnych teorii pola, klasycznych i kwantowych. Kluczowe w tym podejściu pojęcie (oczekiwanej) IF wprowadził Fisher na gruncie własnych rozważań związanych z oszacowywaniem parametrów modeli, podlegających badaniu statystycznemu w ramach ogólnej metody ekstremalnej wartości funkcji wiarygodności L . IF opisuje lokalne własności funkcji wiarygodności $L \equiv P(y|\Theta)$, która formalnie jest łącznym prawdopodobieństwem (lub łączną gęstością prawdopodobieństwa) danych y , lecz jest rozumiana jako funkcja zbioru parametrów Θ , który z kolei tworzy współrzędne w przestrzeni statystycznej. Analiza statystyczna modeli fizycznych idzie jak dotąd dwoma nurtami.

Pierwszy z nich, geometryczny, próbuje opisać metodę statystyczną wprowadzoną w latach 20 poprzedniego wieku przez Fishera [1], twórcę podstaw techniki statystycznej otrzymywania dobrych estymatorów MNW, w jak się okazało naturalnym dla niej środowisku geometrii różniczkowej. Rozwijając MNW, już w 1945 roku C.R. Rao [2] zauważył, że macierz informacji Fishera określa metrykę Riemanną i badał strukturę modeli statystycznych z punktu widzenia geometrii Riemannowskiej. Z kolei B. Efron [3] badając jednoparametrowe modele i analizując ich asymptotyczne własności dla procedur estymacyjnych, wprowadził i odkrył użyteczność pojęcia statystycznej krzywizny. A.P. Dawid [4] wprowadził pojęcie koneksji na przestrzeni wszystkich dodatnio określonych rozkładów prawdopodobieństwa, pokazując, że ze względu na tę koneksję statystyczna krzywizna jest krzywizną zewnętrzną. Jednak problemem Dawida był nieskończony wymiar przestrzeni rozkładów. W roku 1980 S. Amari [5] opublikował systematyczne ujęcie teorii Dawida dla modeli skończone wymiarowych i podał spójne określenie α -koneksji (wprowadzonej wcześniej poza kontekstem statystycznej estymacji przez N.N. Chentsova). W 1982 S. Amari wraz z H. Nagaoka [6] wykazali dualność płaskich przestrzeni modeli eksponencjalnych z e-koneksją i modeli mieszanych z m-koneksją.

Procedury estymacyjne statystycznego opisu mechaniki kwantowej (falowej) poszły dwoma drogami. Pierwsza z nich związana jest z naturalnym dla mechaniki kwantowej formalizmem macierzy gęstości, druga z konstrukcją zasad informacyjnych (entropijnych). W przypadku formalizmu macierzy gęstości, ich zbiór $S = \bigcup_{r=1}^k S_r$ ($S_r \cup S_i = \emptyset$, $i \neq r$) dla przypadku skończone wymiarowych przestrzeni Hilberta \mathcal{H} , tworzy zbiór wypukły. Dla stanów czystych, podzbiór S_1 tego zbioru tworzą punkty ekstremalne, a przestrzeń stanów czystych związana z nim może być utożsamiona z zespoloną przestrzenią rzutową CP^{k-1} , ($k = \dim \mathcal{H}$). Na przestrzeni tej można wprowadzić (z dokładnością do stałej) metrykę Riemannowską nazywaną metryką Fubiniego-Study, która jest kwantową wersją metryki Rao-Fishera. Statystyczną estymacją w modelach dla stanów czystych zajmowali się między innymi Fujiwara, Nagaoka i Matsumoto [7]. Natomiast w przypadku podzbioru S_k zbioru S dualna struktura z metryką może być traktowana jako kwantowy analog metryki Rao-Fishera z $\pm\alpha$ -koneksją.

Drugim nurtem, który wyłonił się w ostatnich kilkunastu latach i którym szedł rozwój zastosowań MNW oraz pojęcia obserwowanej i oczekiwanej IF w fizyce jest formalizm ekstremalnej fizycznej informacji (EFI) opracowany przez Friedena i jego współpracowników, w szczególności Soffera [8]. Konstrukcję modeli fizycznych z wykorzystaniem informacji Fishera zapoczątkował Frieden, podając metodę wyprowadzenia z informacji Fishera członu kinetycznego modeli fizycznych. Następnie zapostulował wprowadzenie dwóch zasad informacyjnych służących do ustalenia związku pomiędzy pojemnością kanału informacyjnego I oraz informacją strukturalną Q , tzn. poprzez zapostulowaną nową strukturalną zasadę informacyjną skonstruował on człony strukturalne rozważanych przez siebie modeli. W odróżnieniu od Friedena stosujemy jednak inne [9], bardziej fizyczne a mniej informacyjne, podejście do konstrukcji podstawowych zasad informacyjnych, posługując się pojęciem całkowitej fizycznej informacji $K = I + Q$, a nie wprowadzonym przez Friedena pojęciem zmiany fizycznej informacji. Różnica ta, chociaż nie powoduje zasadniczo rachunkowych zmian w sposobie wyprowadzenia równań ruchu bądź równań generujących rozkład dla rozważanych do tej pory problemów, jednak zmieniając pojęcie informacji fizycznej oraz jej rozkładu na kinetyczne i strukturalne stopnie swobody, idzie w linii prowadzonych ostatnio badań nad konstrukcją zasady ekwipartycji dla entropii. To inne niż Friedenowskie podejście do pojęcia fizycznej informacji powoduje również zmiany w pojmowaniu istoty przekazu informacji w procesie pomiaru przy jej przekazie od strukturalnych do kinetycznych stopni swobody. Pomimo różnic samą metodę będziemy dalej nazywać podejściem Friedenowskim. Gdyby pominąć chwilowo proces pomiaru, w metodzie Friedena próbkowanie przestrzeni jest wykonywane przez układ, który poprzez właściwe dla niego pole (i związane z nim amplitudy) o randze N będącej wielkością próby, próbkuje jego kinetycznymi (Fisherowskimi) stopniami swobody dostępną mu przestrzeń konfiguracyjną. Następnie, ponieważ IF jest infinitezymalnym typem entropii Kulbacka-Leiblera, to zauważając, że entropia Kulbacka-Leiblera jest wykorzystywana w statystyce do przeprowadzania testów wyboru modeli, pojawia się przypuszczenie, że IF może poprzez narzucenie na nią odpowiednich dodatkowych ograniczeń, zapostulowanych w postaci wspomnianych dwóch zasad informacyjnych, wariacyjnej (skalarnej) oraz strukturalnej (wewnętrznej), doprowadzić do wyprowadzenia równań ruchu bądź równań stanu układów fizycznych, najlepszych z punktu widzenia owych informacyjnych zasad. Na tym zasadza się Friedenowska idea estymacji fizycznych modeli.

Zasady informacyjne mają uniwersalną postać, jednak ich konkretne realizacje zależą od fizyki rozważanego zagadnienia. Pierwsza z zasad informacyjnych, strukturalna, opisuje wewnętrzne charakterystyki układu związane np. z jego spinem. Druga, wariacyjna, prowadzi do otrzymania właściwej relacji dyspersyjnej dla układu. Ciekawą sprawą jest, że wiele rachunków można wykonać dla przypadku, dla którego całkowita fizyczna informacja układu (oraz jej gęstość) dzieli się na dwie równe (lub z czynnikiem $1/2$) części, tzn. pojemność kanału informacyjnego oraz informację strukturalną, mając swoją całkowitą wartość równą zero. Frieden podał informacyjne wyprowadzenie równania Kleina-Gordona dla ogólnego modelu pola z rangą N , z szczególnym uwzględnieniem przypadku pola skalarnego z $N=2$. Dla pola spinorowego z $N=8$ otrzymał równanie Diraca a dla $N=4$ równania Maxwella. Procedura jest na tyle ogólna, że umożliwia opis pól Rarity-Schwingera, ogólnej teorii względności oraz wprowadzenie transformacji cechowania [8]. W oparciu o wprowadzone zasady informacyjne Frieden podał również informacyjne wyprowadzenie zasady nieoznaczoności Heisenberga oparte ze statystycznego punktu widzenia o twierdzenie Rao-Craméra dla informacji Fishera oraz jej relację z pojemnością informacyjną układu zapisaną w reprezentacji pędowej, czyli po dokonaniu transformacji Fouriera. Transformacja Fouriera pełni zresztą w całym formalizmie Friedenowskim rolę wyjątkową, będąc jednym z typów samosplątania wewnątrz przestrzeni konfiguracyjnej

układu, o czym wspomnimy nieco poniżej. Frieden podał również wyprowadzenie klasycznej fizyki statystycznej, tzn. jej podstawowych rozkładów, Boltzmann dla energii oraz Maxwella-Boltzmann dla pędu jak również pewnych rozkładów, które zinterpretował jako odpowiadające przypadkom nierównowagowym. Kolejną sprawą było wyprowadzenie górnego ograniczenia na tempo zmiany entropii układu dla przykładów klasycznego strumienia cząstek, gęstości rozkładu ładunku, czteropotencjału elektrodynamicznego oraz strumienia cząstek o spinie $1/2$ [8]. Podał również opis teorii pomiaru z szumem wykorzystując wariacyjny formalizm EFI pozwalający na opis redukcji funkcji falowej w trakcie pomiaru urządzeniem dającym swój własny szum. Mianowicie po dokonaniu ekstremalizacji sumy informacji fizycznej K niemierzonego układu oraz funkcjonału opisującego własności układu pomiarowego (a będącego splotem funkcji log-wiarygodności dla funkcji przyrządu splecionej nieliniowo z rozkładem układu), otrzymał równanie ruchu, które (po przejściu do nierelatywistycznej granicy Schrödingera) daje równanie typu Feynmana-Mensky'ego z nieliniowym członem opisującym kolaps funkcji falowej w pomiarze. Ciekawe jest to, że w tym przypadku w pełni ujawnia się traktowanie czasu na równi ze zmiennymi przestrzennymi, czyli jako zmiennej losowej z rozkładem prawdopodobieństwa. Przedstawiona w skrypcie postać zasad informacyjnych [10] daje formalnie te same równania ewolucji funkcji falowej układu oplecionej funkcją pomiarową przyrządu, jednak otrzymana interpretacja jest zdecydowanie bardziej spójna niż Friedenowska, pozwalając na jednoznaczne rozróżnienie układu poza pomiarem od układu w pomiarze.

Przedstawione w skrypcie, fundamentalna postać *obserwowanej* strukturalnej zasady informacyjnej oraz jej postać *oczekiwana*, [9, 10], wykorzystywane za Friedenem dla każdego omawianego problemu, zostały ostatnio wyprowadzone dla wartości tzw. współczynnika efektywności $\kappa = 1$ [10]. Zasada strukturalna sugeruje splątanie przestrzeni danych obserwowanych z nieobserwowaną przestrzenią konfiguracyjną układu [10]. Zatem informacja strukturalna Q [10] reprezentuje również informację o splątaniu widocznym w korelacji danych w przestrzeni pomiarowej, a EFI może być wykorzystywana jako mechanizm w estymacji stanów splątanych. Np. w przypadku problemu EPR-Bohma, splątanie zachodzi pomiędzy rzutem spinu obserwowanej cząstki i nieobserwowaną konfiguracją łączną układu, a w przypadku relatywistycznych równań ruchu otrzymujemy splątanie kinetycznych i strukturalnych (masa) stopni swobody, czego wyrazem jest związek stanu obserwowanej cząstki w czasoprzestrzeni z jej własnym stanem w przestrzeni energetyczno-pędowej. Ten drugi przypadek jest przykładem wspomnianego samosplątania opisanego transformatą Fouriera. Ponieważ Q związane jest tu z masą cząstki, zatem w podejściu informacyjnym można wyciągnąć również wniosek, że samosplątanie powinno pomóc w odczytaniu struktury wewnętrznej cząstek. W końcu pojęcie informacji Fishera i jej reinterpretacja przez Friedena jako członu kinetycznego teorii, pozwoliła na przeprowadzenie informacyjnego dowodu [9] o niewyprowadzalności mechaniki kwantowej (falowej) oraz każdej teorii pola, dla której ranga pola N jest skończona, z mechaniki klasycznej.

Temat skryptu dotyczy więc fundamentalnego zagadnienia związanego z określeniem statystycznej procedury estymacji modeli fizycznych. Jego realizacja wymaga znajomości problemów związanych z stosowaniem statystycznej MNW oraz fizycznej EFI dla konstrukcji modeli fizycznych, jak również podstaw metod geometrii różniczkowej.

Na koniec uwaga słownikowa i podsumowanie treści metody EFI. Pojęcie “likelihood function” zostało wprowadzony przez Fishera jako mające związek z prawdopodobieństwem. Również słownikowo powinno być ono przetłumaczone jako “funkcja możliwości”. Zastosowano jednak tłumaczenie “funkcja wiarygodności”. Jako posumowanie istoty przedstawionej metody, powiedzmy, że jest ona wyrazem

zastosowania informacji Fishera w teorii pola w ujęciu Friedena, którego inspiracja pochodzi z obszaru optyki.

Temat skryptu związany jest z dociekaniem, które dane mi było prowadzić wspólnie ze Sławomirem Manią, Dorotą Mroziakiewicz, Janem Ślaskowskim, Robertem Szafronem i Sebastianem Zającem, którym za te dociekania i rozmowy dziękuję.

Dziękuję mojej żonie Grażynie za uważne przeczytanie tekstu skryptu.

Rozdział 1

Metoda największej wiarygodności

Z powodu możliwości zastosowania *metody największej wiarygodności* (MNW) do rozwiązania wielu, bardzo różnych problemów estymacyjnych, stała się ona obecnie zarówno metodą podstawową jak również punktem wyjścia dla różnych metod analizy statystycznej. Jej wszechstronność związana jest, po pierwsze z możliwością przeprowadzenia analizy statystycznej dla małej próbki, opisu zjawisk nieliniowych oraz zastosowania zmiennych losowych posiadających zasadniczo dowolny *rozkład prawdopodobieństwa* [11], oraz po drugie, szczególnymi własnościami otrzymywanych przez nią estymatorów, które okazują się być zgodne, asymptotycznie nieobciążone, efektywne oraz dostateczne [11]. MNW zasadza się na intuicyjnie jasnym postulatcie przyjęcia za prawdziwe takich wartości parametrów rozkładu prawdopodobieństwa zmiennej losowej, które maksymalizują funkcję wiarygodności realizacji konkretnej próbki.

1.1 Podstawowe pojęcia MNW

Rozważmy zmienną losową Y [11], która przyjmuje wartości y zgodnie z rozkładem prawdopodobieństwa $p(y|\theta)$, gdzie $\theta = (\vartheta_1, \vartheta_2, \dots, \vartheta_k)^T \equiv (\vartheta_s)_{s=1}^k$, jest zbiorem k parametrów tego rozkładu (T oznacza transpozycję). Zbiór wszystkich możliwych wartości y zmiennej Y oznaczmy przez \mathcal{Y} . Gdy $k > 1$ wtedy θ nazywamy parametrem *wektorowym*. W szczególnym przypadku $k = 1$ mamy $\theta = \vartheta$. Mówimy wtedy, że parametr θ jest parametrem *skalarnym*.

Pojęcie próby i próbki: Rozważmy *zbiór danych* $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N$ otrzymanych w N obserwacjach zmiennej losowej Y .

Każda z danych \mathbf{y}_n , $n = 1, 2, \dots, N$, jest generowana z rozkładu $p_n(\mathbf{y}_n|\theta_n)$ zmiennej losowej Y w populacji, którą charakteryzuje wartość parametru wektorowego $\theta_n = (\vartheta_1, \vartheta_2, \dots, \vartheta_k)_n^T \equiv ((\vartheta_s)_{s=1}^k)_n$, $n = 1, 2, \dots, N$. Stąd zmienną Y w n -tej populacji oznaczmy Y_n . Zbiór zmiennych losowych $\tilde{Y} = (Y_1, Y_2, \dots, Y_N) \equiv (Y_n)_{n=1}^N$ nazywamy N -wymiarową *próbą*.

Konkretną realizację $y = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N) \equiv (\mathbf{y}_n)_{n=1}^N$ próby \tilde{Y} nazywamy *próbką*. Zbiór wszystkich możliwych realizacji y próby \tilde{Y} tworzy przestrzeń próby (układu) oznaczaną jako \mathcal{B} .

Określenie: Ze względu na to, że n jest indeksem konkretnego punktu pomiarowego próby, rozkład $p_n(\mathbf{y}_n|\theta_n)$ będziemy nazywali rozkładem *punktowym* (czego nie należy mylić z np. rozkładem dyskretnym).

Określenie funkcji wiarygodności: Centralnym pojęciem MNW jest *funkcja wiarygodności* $L(y; \Theta)$ (pojawienia się) próbki $y = (\mathbf{y}_n)_{n=1}^N$, nazywana też *wiarygodnością próbki*. Jest ona funkcją parametru Θ .

Przez wzgląd na zapis stosowany w fizyce, będziemy stosowali oznaczenie $P(y|\Theta) \equiv L(y; \Theta)$, które podkreśla, że formalnie *funkcja wiarygodności jest łącznym rozkładem prawdopodobieństwa*¹ pojawienia się realizacji $y \equiv (\mathbf{y}_n)_{n=1}^N$ próby $\tilde{Y} \equiv (Y_n)_{n=1}^N$, to znaczy:

$$P(\Theta) \equiv P(y|\Theta) = \prod_{n=1}^N p_n(\mathbf{y}_n|\theta_n). \quad (1.3)$$

Zwrócenie uwagi w (1.3) na występowanie y w argumencie funkcji wiarygodności oznacza, że może być ona rozumiana jako statystyka $P(\tilde{Y}|\Theta)$. Z kolei skrócone oznaczenie $P(\Theta)$ podkreśla, że centralną sprawą w MNW jest fakt, że funkcja wiarygodności jest funkcją nieznanych parametrów:

$$\Theta = (\theta_1, \theta_2, \dots, \theta_N)^T \equiv (\theta_n)_{n=1}^N \quad \text{przy czym} \quad \theta_n = (\vartheta_{1n}, \vartheta_{2n}, \dots, \vartheta_{kn})^T \equiv ((\vartheta_s)_{s=1}^k)_n, \quad (1.4)$$

gdzie θ_n jest wektorowym parametrem populacji określonej przez indeks próby n . W toku analizy chcemy oszacować wektorowy parametr Θ .

Zbiór wartości parametrów $\Theta = (\theta_n)_{n=1}^N$ tworzy współrzędne rozkładu prawdopodobieństwa rozumianego jako punkt w $d = k \times N$ - wymiarowej (podprzestrzeni) przestrzeni statystycznej \mathcal{S} [6]. Temat ten rozwinieśmy w Rozdziale 2.2.

Uwaga o postaci rozkładów punktowych: W skrypcie zakładamy, że “punktowe” rozkłady $p_n(\mathbf{y}_n|\theta_n)$ dla poszczególnych pomiarów n w N elementowej próbie są *niezależne*².

W ogólności w treści skryptu, rozkłady punktowe $p_n(\mathbf{y}_n|\theta_n)$ zmiennych Y_n chociaż są *tego samego typu*, jednak nie spełniają warunku (1.5) charakterystycznego dla próby prostej. Taka ogólna sytuacja ma np. miejsce w analizie regresji (Rozdział 1.3).

¹**Miara produktowa:** Niech będzie danych N przestrzeni probabilistycznych $\{\Omega_1, \mathcal{F}_1, P_1\}, \dots, \{\Omega_N, \mathcal{F}_N, P_N\}$, gdzie $\Omega_n, \mathcal{F}_n, P_n$ są, dla każdego $n = 1, 2, \dots, N$, odpowiednio n -tą przestrzenią zdarzeń, σ - ciałem na Ω_n oraz miarą probabilistyczną. Wprowadźmy na produkcie $\Omega = \Omega_1 \times \dots \times \Omega_N$ tzw. σ - ciało produktowe $\mathcal{F} \equiv \mathcal{F}_1 \otimes \dots \otimes \mathcal{F}_N$ będące najmniejszym σ - ciałem zawierającym zbiory postaci $A_1 \times \dots \times A_N$ gdzie $A_1 \in \mathcal{F}_1, \dots, A_N \in \mathcal{F}_N$.

Na produkcie Ω można zdefiniować miarę produktową P taką, że:

$$P(A_1 \times A_2 \times \dots \times A_N) = P_1(A_1) P_2(A_2) \dots P_N(A_N). \quad (1.1)$$

Zmienne niezależne: Zmienne losowe Y_1, \dots, Y_N o wartościach w \mathbb{R} określone odpowiednio na $\{\Omega_1, \mathcal{F}_1, P_1\}, \dots, \{\Omega_N, \mathcal{F}_N, P_N\}$ nazywamy *niezależnymi*, gdy dla każdego ciągu zbiorów borelowskich B_1, \dots, B_N zachodzi równość:

$$P(Y_1 \in B_1, Y_2 \in B_2, \dots, Y_N \in B_N) = P_1(Y_1 \in B_1) P_2(Y_2 \in B_2) \dots P_N(Y_N \in B_N). \quad (1.2)$$

Ponieważ po prawej stronie (1.2) stoją zdarzenia losowe $Y_n^{-1}(B_n)$, gdzie B_n należy do σ - ciała $\mathbb{B}(\mathbb{R})$, zatem stwierdzenie (1.2) oznacza, że *zmienne losowe są niezależne wtedy i tylko wtedy, gdy σ - ciała \mathcal{F}_n generowane przez zmienne losowe $Y_n, n = 1, 2, \dots, N$, są niezależne* [12].

²W przypadku analizy jednej zmiennej losowej Y , rozkłady te obok niezależności spełniają dodatkowo warunek:

$$p_n(\mathbf{y}_n|\theta_n) = p(y|\theta) , \quad (1.5)$$

co oznacza, że próba jest *prosta*.

Pojęcie estymatora parametru: Załóżmy, że dane $y = (y_n)_{n=1}^N$ są generowane losowo z punktowych rozkładów prawdopodobieństwa $p_n(y_n|\theta_n)$, $n = 1, 2, \dots, N$, które chociaż nie są znane, to jednak założono o nich, że dla każdego n należą do określonej, tej samej klasy modeli. Zatem funkcja wiarygodności (1.3) należy do określonej, $d = k \times N$ - wymiarowej, przestrzeni statystycznej \mathcal{S} .

Celem analizy jest oszacowanie nieznanego parametru Θ , (1.4), poprzez funkcję:

$$\hat{\Theta} \equiv \hat{\Theta}(\tilde{Y}) = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_N)^T \equiv (\hat{\theta}_n)_{n=1}^N \quad \text{gdzie} \quad \hat{\theta}_n = (\hat{\vartheta}_{1n}, \hat{\vartheta}_{2n}, \dots, \hat{\vartheta}_{kn})^T \equiv ((\hat{\vartheta}_s)_{s=1}^k)_n, \quad (1.6)$$

mającą $d = k \times N$ składowych.

Każda z funkcji $\hat{\vartheta}_{kn} \equiv \hat{\vartheta}_{kn}(\tilde{Y})$ jako funkcja próby jest *statystyką*, którą przez wzgląd na to, że służy do oszacowywania wartości parametru ϑ_{kn} nazywamy estymatorem tego parametru. *Estymator parametru nie może zależeć od parametru, który oszacowuje*³.

Podsumowując, odwzorowanie:

$$\hat{\Theta} : \mathcal{B} \rightarrow \mathbf{R}^d, \quad (1.7)$$

gdzie \mathcal{B} jest przestrzenią próby, jest estymatorem parametru (wektorowego) Θ .

Równania wiarygodności: Będąc funkcją $\Theta = (\theta_n)_{n=1}^N$, funkcja wiarygodności służy do konstrukcji estymatorów $\hat{\Theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_N)^T \equiv (\hat{\theta}_n)_{n=1}^N$ parametrów $\Theta \equiv (\theta_n)_{n=1}^N$. Procedura polega na wyborze takich $(\hat{\theta}_n)_{n=1}^N$, dla których funkcja wiarygodności przyjmuje maksymalną wartość, skąd statystyki te nazywamy estymatorami MNW.

Zatem warunek konieczny otrzymania estymatorów $\hat{\Theta}$ MNW sprowadza się do znalezienia rozwiązania układu $d = k \times N$ tzw. *równań wiarygodności* [1]:

$$S(\Theta)|_{\Theta=\hat{\Theta}} \equiv \frac{\partial}{\partial \Theta} \ln P(y|\Theta)|_{\Theta=\hat{\Theta}} = 0, \quad (1.8)$$

gdzie zagadnienie maksymalizacji funkcji wiarygodności $P(y|\Theta)$ sprowadzono do (na ogół) analitycznie równoważnego mu problemu maksymalizacji jej logarytmu $\ln P(y|\Theta)$.

Określenie funkcji wynikowej: Funkcję $S(\Theta)$ będącą gradientem logarytmu funkcji wiarygodności:

$$S(\Theta) \equiv \frac{\partial}{\partial \Theta} \ln P(y|\Theta) = \begin{pmatrix} \frac{\partial \ln P(y|\Theta)}{\partial \theta_1} \\ \vdots \\ \frac{\partial \ln P(y|\Theta)}{\partial \theta_N} \end{pmatrix} \quad \text{gdzie} \quad \frac{\partial \ln P(y|\Theta)}{\partial \theta_n} = \begin{pmatrix} \frac{\partial \ln P(y|\Theta)}{\partial \vartheta_{1n}} \\ \vdots \\ \frac{\partial \ln P(y|\Theta)}{\partial \vartheta_{kn}} \end{pmatrix}, \quad (1.9)$$

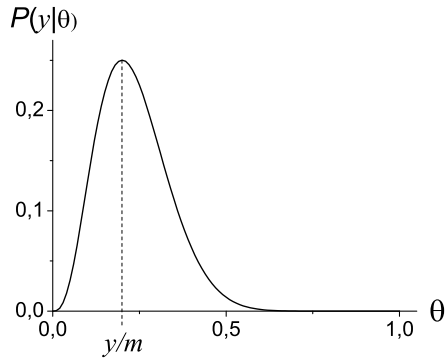
nazywamy *funkcją wynikową*.

Po otrzymaniu (wektora) estymatorów $\hat{\Theta}$, *zmaksymalizowaną* wartość funkcji wiarygodności definiujemy jako numeryczną wartość funkcji wiarygodności powstałą przez podstawienie do $P(y|\Theta)$ wartości oszacowanej $\hat{\Theta}$ w miejsce parametru Θ .

Przykład: Rozważmy problem estymacji skalarne parametru, tzn. $\Theta = \theta$ (tzn. $k = 1$ oraz $N = 1$), dla zmiennej losowej Y opisanej rozkładem dwumianowym (Bernoulliego):

$$P(y|\theta) = \binom{m}{y} \theta^y (1 - \theta)^{m-y}. \quad (1.10)$$

³Natomiast rozkład estymatora oszacowywanego parametru, zależy od tego parametru.



Rysunek 1.1: Graficzna ilustracja metody największej wiarygodności dla $P(y|\theta)$ określonego wzorem (1.10) dla rozkładu dwumianowego. Przyjęto wartość parametru $m = 5$. Na skutek pomiaru zaobserwowano wartość Y równą $y = 1$. Maksimum $P(y|\theta)$ przypada na wartość θ równą punktowemu oszacowaniu $\hat{\theta} = y/m = 1/5$ tego parametru. Maksymalizowana wartość funkcji wiarygodności wynosi $P(y|\hat{\theta})$.

Estymacji parametru θ dokonamy na podstawie *pojedynczej* obserwacji (długość próby $N = 1$) zmiennej Y , której iloraz Y/m nazywamy *częstością*. Parametr m charakteryzuje rozkład zmiennej Bernoulliego Y (i nie ma związku z długością N próby).

Zatem ponieważ $y \equiv (y_1)$, więc $P(y|\theta)$ jest funkcją wiarygodności dla $N = 1$ wymiarowej próby. Jej logarytm wynosi:

$$\ln P(y|\theta) = \ln \binom{m}{y} + y \ln \theta + (m - y) \ln (1 - \theta) . \quad (1.11)$$

W rozważanym przypadku otrzymujemy jedno równanie wiarygodności (1.8):

$$S(\theta) = \frac{1}{\theta} y - \frac{1}{1 - \theta} (m - y) \Big|_{\theta = \hat{\theta}} = 0 \quad (1.12)$$

a jego rozwiązanie daje estymator MNW parametru θ rozkładu dwumianowego, równy:

$$\hat{\theta} = \frac{y}{m} \quad (1.13)$$

Ilustracją powyższej procedury znajdowania wartości estymatora parametru θ jest Rysunek 1.1 (gdzie przyjęto $m = 5$), przy czym na skutek pomiaru zaobserwowano wartość Y równą $y = 1$.

1.2 Wnioskowanie w MNW

Z powyższych rozważań wynika, że konstrukcja punktowego oszacowania parametru w MNW oparta jest o postulat maksymalizacji funkcji wiarygodności przedstawiony powyżej. Jest on wstępem do statystycznej procedury wnioskowania. Kolejnym krokiem jest konstrukcja przedziału wiarygodności. Jest on odpowiednikiem przedziału ufności, otrzymywanego w częstotliwościowym podejściu statystyki klasycznej do procedury estymacyjnej. Do jego konstrukcji niezbędna jest znajomość rozkładu prawdopodobieństwa estymatora parametru, co (dzięki “porządnym” granicznym własnościom stosowanych estymatorów) jest możliwe niejednokrotnie jedynie asymptotycznie, tzn. dla wielkości próby dążącej do nieskończoności. Znajomość rozkładu estymatora jest też niezbędna we wnioskowaniu statystycznym odnoszącym się do weryfikacji hipotez.

W sytuacji, gdy nie dysponujemy wystarczającą ilością danych, potrzebnych do przeprowadzenia skutecznego częstotliwościowego wnioskowania, Fisher [13] zaproponował do określenia niepewności dotyczącej parametru Θ wykorzystanie maksymalizowanej wartości funkcji wiarygodności.

Przedział wiarygodności jest zdefiniowany jako zbiór wartości parametru Θ , dla których funkcja wiarygodności osiąga (umownie) wystarczająco wysoką wartość, tzn.:

$$\left\{ \Theta, \frac{P(y|\Theta)}{P(y|\hat{\Theta})} > c \right\}, \quad (1.14)$$

dla pewnego parametru obciążenia c , nazywanego *poziomem wiarygodności*.

Iloraz wiarygodności:

$$P(y|\Theta)/P(y|\hat{\Theta}) \quad (1.15)$$

reprezentuje pewien typ unormowanej wiarygodności i jako taki jest wielkością skalarną. Jednak z powodu niejasnego znaczenia określonej wartości parametru obciążenia c pojęcie to wydaje się być na pierwszy rzut oka za słabe, aby dostarczyć taką precyzję wypowiedzi jaką daje analiza częstotliwościowa.

Istotnie, wartość c nie odnosi się do żadnej wielkości obserwowanej, tzn. na przykład 1%-we ($c = 0,01$) obciążenie nie ma ścisłego probabilistycznego znaczenia. Inaczej ma się sprawa dla częstotliwościowych przedziałów ufności. W tym przypadku wartość współczynnika $\alpha = 0,01$ oznacza, że gdybyśmy rozważyli realizację przedziału ufności na poziomie ufności $1 - \alpha = 0,99$, to przy pobraniu nieskończonej (w praktyce wystarczająco dużej) liczby próbek, 99% wszystkich wyznaczonych przedziałów ufności pokryłoby prawdziwą (teoretyczną) wartość parametru Θ w populacji generalnej (składającej się z N podpopulacji). Pomimo tej słabości MNW zobaczymy, że rozbudowanie analizy stosunku wiarygodności okazuje się być istotne we wnioskowaniu statystycznym analizy doboru modeli i to aż po konstrukcję równań teorii pola.

1.2.1 Wiarygodnościowy przedział ufności

Przykład rozkładu normalnego z jednym estymowanym parametrem: Istnieje przypadek pozwalający na prostą interpretację przedziału wiarygodnościowego jako przedziału ufności. Dotyczy on zmiennej Y posiadającej rozkład Gaussa oraz sytuacji gdy (dla próby prostej) interesuje nas estymacja skalarne parametru θ będącego wartością oczekiwaną $E(Y)$ zmiennej Y . Przypadek ten

omówimy poniżej. W ogólności, przedział wiarygodności posiadający określony poziom ufności jest nazywany przedziałem ufności.

Częstotliwościowe wnioskowanie o nieznanym parametrze θ wymaga określenia rozkładu jego estymatora, co jest zazwyczaj możliwe jedynie granicznie [6]. Podobnie w MNW, o ile to możliwe, korzystamy przy dużych próbkach z twierdzeń granicznych dotyczących rozkładu ilorazu wiarygodności [6, 13]. W przypadku rozkładu normalnego i parametru skalarne okazuje się, że możliwa jest konstrukcja skończonego wymiaru.

Niech więc zmienna Y ma rozkład normalny $N(\theta, \sigma^2)$:

$$p(y|\theta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\theta)^2}{2\sigma^2}\right). \quad (1.16)$$

Rozważmy próbkę $y \equiv (y_1, \dots, y_N)$, która jest realizacją próby prostej \tilde{Y} i założmy, że *wariancja σ^2 jest znana*. Logarytm funkcji wiarygodności dla $N(\theta, \sigma^2)$ ma postać:

$$\ln P(y|\theta) = -\frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=1}^N (y_n - \theta)^2, \quad (1.17)$$

gdzie ze względu na próbę prostą, w argumencie funkcji wiarygodności wpisano w miejsce $\Theta \equiv (\theta)_{n=1}^N$ parametr θ , jedyny który podlega estymacji.

Korzystając z funkcji wiarygodności (1.17) otrzymujemy oszacowanie MNW parametru θ równe⁴ $\hat{\theta} = \bar{y} = \frac{1}{N} \sum_{n=1}^N y_n$, co pozwala na zapisanie równości $\sum_{n=1}^N (y_n - \theta)^2 = \sum_{n=1}^N ((y_n - \hat{\theta}) + (\hat{\theta} - \theta))^2 = \sum_{n=1}^N (y_n - \hat{\theta})^2 + \sum_{n=1}^N (\hat{\theta} - \theta)^2$. W końcu, nieskomplikowane przekształcenia prowadzą do następującej postaci logarytmu ilorazu wiarygodności (1.15):

$$\ln \frac{P(y|\theta)}{P(y|\hat{\theta})} = -\frac{N}{2\sigma^2} (\hat{\theta} - \theta)^2. \quad (1.21)$$

Statystyka Wilka: Widać, że po prawej stronie (1.21) otrzymaliśmy wyrażenie kwadratowe. Ponieważ \bar{Y} jest nieobciążonym estymatorem parametru θ , co oznacza, że wartość oczekiwana $E(\bar{Y}) = \theta$, zatem (dla rozkładu $Y \sim N(\theta, \sigma^2)$) średnia arytmetyczna \bar{Y} ma rozkład normalny $N\left(\theta, \frac{\sigma^2}{N}\right)$.

⁴ **Postać estymatora parametru skalarnego θ rozkładu $N(\theta, \sigma^2)$:** Korzystając z równania wiarygodności (1.8) dla przypadku skalarne parametru θ , otrzymujemy:

$$S(\theta)|_{\theta=\hat{\theta}} \equiv \frac{\partial}{\partial \theta} \ln P(y|\theta)|_{\theta=\hat{\theta}} = 0, \quad (1.18)$$

skąd dla log-funkcji wiarygodności (1.17), otrzymujemy:

$$\hat{\theta} = \bar{y} = \frac{1}{N} \sum_{n=1}^N y_n. \quad (1.19)$$

Zatem estymatorem parametru θ jest średnia arytmetyczna:

$$\hat{\theta} = \bar{Y} = \frac{1}{N} \sum_{n=1}^N Y_n. \quad (1.20)$$

Estymator i jego realizowaną wartość będziemy oznaczali tak samo, tzn. $\hat{\theta}$ dla przypadku skalarne i $\hat{\Theta}$ dla wektorowego.

Z normalności rozkładu \tilde{Y} wynika, że tzw. *statystyka ilorazu wiarygodności Wilka*:

$$W \equiv 2 \ln \frac{P(\tilde{Y}|\hat{\theta})}{P(\tilde{Y}|\theta)} \sim \chi_1^2, \quad (1.22)$$

ma rozkład χ^2 , w tym przypadku z jednym stopniem swobody [13].

Wyskalowanie statystyki Wilka w przypadku normalnym: Wykorzystując (1.22) możemy wykonać wyskalowanie wiarygodności oparte o możliwość powiązania przedziału wiarygodności z jego częstotliwościowym odpowiednikiem.

Mianowicie z (1.22) otrzymujemy, że dla ustalonego (choć nieznanego) parametru θ prawdopodobieństwo, że iloraz wiarygodności znajduje się w wyznaczonym dla parametru obciążenia c , wiarygodnościowym przedziale ufności, wynosi:

$$P\left(\frac{P(\tilde{Y}|\theta)}{P(\tilde{Y}|\hat{\theta})} > c\right) = P\left(2 \ln \frac{P(\tilde{Y}|\hat{\theta})}{P(\tilde{Y}|\theta)} < -2 \ln c\right) = P(\chi_1^2 < -2 \ln c). \quad (1.23)$$

Zatem jeśli dla jakiegoś $0 < \alpha < 1$ wybierzemy parametr obciążenia:

$$c = e^{-\frac{1}{2}\chi_{1,(1-\alpha)}^2}, \quad (1.24)$$

gdzie $\chi_{1,(1-\alpha)}^2$ jest kwantylem rzędu $100(1-\alpha)\%$ rozkładu χ -kwadrat, to spełnienie przez θ związku:

$$P\left(\frac{P(\tilde{Y}|\theta)}{P(\tilde{Y}|\hat{\theta})} > c\right) = P(\chi_1^2 < \chi_{1,(1-\alpha)}^2) = 1 - \alpha \quad (1.25)$$

oznacza, że przyjęcie wartości c zgodnej z (1.24) daje zbiór możliwych wartości parametru θ :

$$\left\{ \theta, \frac{P(\tilde{Y}|\theta)}{P(\tilde{Y}|\hat{\theta})} > c \right\}, \quad (1.26)$$

nazywany $100(1-\alpha)\%$ -owym *wiarygodnościowym przedziałem ufności*. Jest on odpowiednikiem wyznaczonego na poziomie ufności $(1-\alpha)$ częstotliwościowego przedziału ufności dla θ . Dla analizowanego przypadku rozkładu normalnego z estymacją skalarnego parametru θ oczekiwanego poziomu zjawiska, otrzymujemy po skorzystaniu z wzoru (1.24) wartość parametru obciążenia równego $c = 0.15$ lub $c = 0.04$ dla odpowiednio 95%-owego ($1-\alpha = 0.95$) bądź 99%-owego ($1-\alpha = 0.99$) przedziału ufności. Tak więc w przypadku, *gdy przedział wiarygodności da się wyskalować rozkładem prawdopodobieństwa, parametr obciążenia c posiada własność wielkości obserwowanej interpretowanej częstotliwościowo poprzez związek z poziomem ufności*.

Zwróćmy uwagę, że chociaż konstrukcje częstotliwościowego i wiarygodnościowego przedziału ufności są różne, to *ich losowość wynika* w obu przypadkach z *rozkładu prawdopodobieństwa estymatora $\hat{\theta}$* .

Ćwiczenie: W oparciu o powyższe rozważania wyznaczyć, korzystając z (1.21) ogólną postać przedziału wiarygodności dla skalarnego parametru θ rozkładu normalnego.

1.2.2 Rozkłady regularne

Dla zmiennych o innym rozkładzie niż rozkład normalny, statystyka Wilka W ma w ogólności inny rozkład niż χ^2 [13]. Jeśli więc zmienne nie mają dokładnie rozkładu normalnego lub dysponujemy za małą próbką by móc odwoływać się do (wynikających z twierdzeń granicznych) rozkładów granicznych dla estymatorów parametrów, wtedy związek (1.22) (więc i (1.24)) daje jedynie przybliżone wyskalowanie przedziału wiarygodności rozkładem χ^2 .

Jednakże w przypadkach wystarczająco *regularnych rozkładów*, zdefiniowanych jako takie, w których możemy zastosować przybliżenie kwadratowe:

$$\ln \frac{P(y|\theta)}{P(y|\hat{\theta})} \approx -\frac{1}{2} \text{iF}(\hat{\theta}) (\hat{\theta} - \theta)^2, \quad (1.27)$$

powyższe rozumowanie oparte o wyskalowanie wiarygodności rozkładem χ^2_1 jest w przybliżeniu słuszne. Wielkość $\text{iF}(\hat{\theta})$, która pojawiła się powyżej jest *obserwowaną* informacją Fishera, a powyższa formuła stanowi poważne narzędzie w analizie doboru modeli. Można powiedzieć, że cały skrypt koncentruje się na analizie zastosowania (wartości oczekiwanej) tego wyrażenia i jego uogólnień. Do sprawy tej wrócimy dalej.

Przykład: Rozważmy przypadek parametru skalarne θ w jednym eksperymencie ($N = 1$) ze zmienną Y posiadającą rozkład Bernoulliego z $m = 15$. W wyniku pomiaru zaobserwowaliśmy wartość $Y = y = 3$. Prosta analiza pozwala wyznaczyć wiarygodnościowy przedział ufności dla parametru θ . Ponieważ przestrzeń V_θ parametru θ wynosi $V_\theta = (0, 1)$, zatem łatwo pokazać, że dla $c = 0,01$, $c = 0,1$ oraz $c = 0,5$ miałyby on realizację odpowiednio $(0,019; 0,583)$, $(0,046; 0,465)$ oraz $(0,098; 0,337)$. Widać, że wraz ze wzrostem wartości c , przedział wiarygodności zacieśnia się wokół wartości oszacowania punktowego $\hat{\theta} = y/m = 1/5$ parametru θ i nic dziwnego, bo wzrost wartości c oznacza akceptowanie jako możliwych do przyjęcia tylko takich *modelowych wartości parametru* θ , które gwarantują wystarczająco wysoką wiarygodność próbki.

Powyższy przykład pozwala nabyć pewnej intuicji co do sensu stosowania ilorazu funkcji wiarygodności. Mianowicie po otrzymaniu w pomiarze określonej wartości y/m oszacowującej parametr θ , jesteśmy skłonni preferować model z taką wartością parametru θ , która daje większą wartość (logarytmu) ilorazu wiarygodności $P(y|\theta)/P(y|\hat{\theta})$. Zgodnie z podejściem statystyki klasycznej *nie oznacza to jednak*, że uważamy, że parametr θ ma jakiś rozkład. Jedynie wobec niewiedzy co do modelowej (populacyjnej) wartości parametru θ preferujemy ten model, który daje większą wartość ilorazu wiarygodności w próbce.

1.2.3 Weryfikacja hipotez z wykorzystaniem ilorazu wiarygodności

Powyżej wykorzystaliśmy funkcję wiarygodności do *estymacji wartości parametru* Θ . Funkcję wiarygodności można również wykorzystać w drugim typie wnioskowania statystycznego, tzn. w *weryfikacji hipotez statystycznych*.

Rozważmy prostą hipotezę zerową $H_0 : \Theta = \Theta_0$ wobec złożonej hipotezy alternatywnej $H_1 : \Theta \neq \Theta_0$. W celu przeprowadzenia *testu statystycznego* wprowadźmy unormowaną funkcję wiarygodności:

$$\frac{P(y|\Theta_0)}{P(y|\hat{\Theta})}, \quad (1.28)$$

skonstruowaną przy założeniu prawdziwości hipotezy zerowej. Hipotezę zerową H_0 odrzucamy na rzecz hipotezy alternatywnej, jeśli jej wiarygodność $P(y|\Theta_0)$ jest “za mała”. Sugerowałoby to, że złożona hipoteza alternatywna H_1 zawiera pewną hipotezę prostą, która jest lepiej poparta przez dane otrzymane w próbie, niż hipoteza zerowa.

Jak o tym wspomnieliśmy powyżej, np. 5%-owe obcięcie c w zagadnieniu estymacyjnym, samo w sobie nie mówi nic o frakcji liczby przedziałów wiarygodności pokrywających nieznaną wartość szacowanego parametru. Potrzebne jest wyskalowanie ilorazu wiarygodności. Również dla weryfikacji hipotez skalowanie wiarygodności jest istotne. Stwierdziliśmy, że takie skalowanie jest możliwe wtedy gdy mamy do czynienia z jednoparametrowym przypadkiem rozkładu Gaussa, a przynajmniej z przypadkiem wystarczająco regularnym.

Empiryczny poziom istotności: W przypadku jednoparametrowego, regularnego problemu z $(\Theta \equiv (\theta)_{n=1}^N)$ jak w Przykładzie z Rozdziału 1.2.1, skalowanie poprzez wykorzystanie statystyki Wilka służy otrzymaniu empirycznego poziomu istotności p . Ze związku (1.22) otrzymujemy wtedy przybliżony (a dokładny dla rozkładu normalnego) *empiryczny poziom istotności*:

$$\begin{aligned} p &\approx P\left(\frac{P(\tilde{Y}|\hat{\theta})}{P(\tilde{Y}|\theta_0)} \geq \frac{P(y|\hat{\theta}_{obs})}{P(y|\theta_0)}\right) = P\left(2\ln \frac{P(\tilde{Y}|\hat{\theta})}{P(\tilde{Y}|\theta_0)} \geq -2\ln c_{obs}\right) \\ &= P(\chi_1^2 \geq -2\ln c_{obs}) \text{ , gdzie } c_{obs} \equiv \frac{P(y|\theta_0)}{P(y|\hat{\theta}_{obs})} \text{ ,} \end{aligned} \quad (1.29)$$

przy czym $\hat{\theta}_{obs}$ jest wartością estymatora MNW $\hat{\theta}$ wyznaczoną w obserwowanej (obs) próbie y . Powyższe określenie empirycznego poziomu istotności p oznacza, że w przypadku wystarczająco regularnego problemu [13], istnieje typowy związek pomiędzy prawdopodobieństwem (1.25), a empirycznym poziomem istotności p , podobny do związku jaki istnieje pomiędzy poziomem ufności $1 - \alpha$, a poziomem istotności α w analizie częstotliwościowej. I tak, np. w przypadku jednoparametrowego rozkładu normalnego możemy wykorzystać wartość empirycznego poziomu istotności p do stwierdzenia, że gdy $p \leq \alpha$ to hipotezę H_0 odrzucamy na rzecz hipotezy H_1 , a w przypadku $p > \alpha$ nie mamy podstawy do odrzucenia H_0 .

Problem błędu pierwszego i drugiego rodzaju: Jednakże podobne skalowanie ilorazu wiarygodności okazuje się być znacznie trudniejsze już chociażby tylko w przypadku dwuparametrowego rozkładu normalnego, gdy obok θ estymujemy σ^2 [13]. Wtedy określenie co oznacza sformułowanie „zbyt mała” wartość c jest dość dowolne i zależy od rozważanego problemu lub wcześniejszej wiedzy wynikającej z innych źródeł niż prowadzone statystyczne wnioskowanie. Wybór dużego parametru obcicia c spowoduje, że istnieje większe prawdopodobieństwo popełnienia *błędu pierwszego rodzaju* polegającego na odrzuceniu hipotezy zerowej w przypadku, gdy jest ona prawdziwa. Wybór małego c spowoduje zwiększenie prawdopodobieństwa popełnienia *błędu drugiego rodzaju*, tzn. przyjęcia hipotezy zerowej w sytuacji, gdy jest ona błędna.

1.3 MNW w analizie regresji

Analiza zawarta w całym Rozdziale 1.3 oparta jest na przedstawieniu metody MNW w analizie regresji klasycznej podanym w [14] i [15].

W metodzie regresji klasycznej, estymatory parametrów strukturalnych modelu regresji są otrzymywane arytmetyczną metodą najmniejszych kwadratów (MNK). Zmienne objaśniające $X_n = x_n$, $n = 1, \dots, N$, nie mają wtedy charakteru stochastycznego, co oznacza, że eksperyment jest ze względu na nie kontrolowany.

MNK polega na minimalizacji sumy kwadratów odchyłek obserwowanych wartości zmiennej objaśnianej (tzw. odpowiedzi) od ich wartości teoretycznych spełniających równanie regresji. MNK ma znaczenie probabilistyczne tylko w przypadku analizy standardowej, gdy zmienna objaśniana Y ma rozkład normalny. Jej estymatory pokrywają się wtedy z estymatorami MNW. Pokażemy, że tak się sprawy mają.

Założmy, że zmienne Y_1, Y_2, \dots, Y_N odpowiadające kolejnym wartościom zmiennej objaśniającej, x_1, x_2, \dots, x_N , są względem siebie niezależne i mają rozkład normalny ze średnią $\mu_n = E(Y|x_n) = E(Y_n)$ zależną od wariantu zmiennej objaśniającej x_n , oraz taką samą wariancję $\sigma^2(Y_n) = \sigma^2(Y)$. Funkcja wiarygodności próbki (y_1, y_2, \dots, y_N) dla normalnego klasycznego modelu regresji z parametrem $\Theta = \mu \equiv (\mu_n)_{n=1}^N$, ma postać:

$$\begin{aligned} P(\mu) \equiv P(y|\mu) &= \prod_{n=1}^N f(y_n|\mu_n) = \prod_{n=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2} (y_n - \mu_n)^2\right\} \\ &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{n=1}^N (y_n - \mu_n)^2\right\}, \end{aligned} \quad (1.30)$$

gdzie $f(y_n|\mu_n)$, $n = 1, 2, \dots, N$, są punktowymi rozkładami gęstości prawdopodobieństwa Gaussa. Widać, że maksymalizacja $P(\mu)$ ze względu na $(\mu_n)_{n=1}^N$ pociąga za sobą minimalizację sumy kwadratów reszt⁵ (SKR):

$$SKR = \sum_{n=1}^N (y_n - \mu_n)^2, \quad (1.31)$$

gdzie $\mu_n = E(Y|x_n)$ jest postulowanym modelem regresji. Zatem w standardowej, klasycznej analizie regresji, estymatory MNW pokrywają się z estymatorami MNK. Widać, że procedura minimalizacji dla SKR prowadzi do liniowej w Y_n postaci estymatorów $\hat{\mu}_n$ parametrów μ_n .

Problem z nieliniowym układem równań wiarygodności: Jednak rozwiązanie układu równań wiarygodności (1.8) jest zazwyczaj nietrywialne. Jest tak, gdy otrzymany w wyniku ekstremizacji układ algebraicznych równań wiarygodności dla estymatorów jest nieliniowy, co w konsekwencji oznacza, że możemy nie otrzymać ich w zwartej analitycznej postaci. Przykładem może być analiza regresji Poissona, w której do rozwiązania równań wiarygodności wykorzystujemy metody iteracyjne. W takich sytuacjach wykorzystujemy na ogół jakiś program komputerowy do analizy statystycznej, np. zawarty w pakiecie SAS. Po podaniu postaci funkcji wiarygodności, program komputerowy dokonuje

⁵SSE w literaturze angielskiej.

jej maksymalizacji rozwiązując układ (1.8) np. metodą Newton-Raphson’a [13, 15], wyznaczając numerycznie wartości estymatorów parametrów modelu.

Testy statystyczne: Logarytm ilorazu wiarygodności jest również wykorzystywany w analizie regresji do przeprowadzania testów statystycznych przy weryfikacji hipotez o nie występowaniu braku dopasowania modelu mniej złożonego, tzw. “niższego”, o mniejszej liczbie parametrów, w stosunku do bardziej złożonego modelu “wyższego”, posiadającego większą liczbę parametrów. Statystyka wykorzystywana do tego typu testów ma postać [14, 13, 16]:

$$-2 \ln \frac{P(\tilde{Y}|\hat{\Theta}_1)}{P(\tilde{Y}|\hat{\Theta}_2)} \quad (1.32)$$

gdzie $P(\tilde{Y}|\hat{\Theta}_1)$ jest maksymalizowaną wartością funkcji wiarygodności dla modelu mniej złożonego, a $P(\tilde{Y}|\hat{\Theta}_2)$ dla modelu bardziej złożonego. Przy prawdziwości hipotezy zerowej H_0 o braku konieczności rozszerzania modelu niższego do wyższego, statystyka (1.32) ma asymptotycznie rozkład χ^2 z liczbą stopni swobody równą różnicy liczby parametrów modelu wyższego i niższego.

Analogia współczynnika determinacji: Maksymalizowana wartość funkcji wiarygodności zachowuje się podobnie jak *współczynnik determinacji* R^2 [14, 15], tzn. rośnie wraz ze wzrostem liczby parametrów w modelu, zatem wielkość pod logarytmem należy do przedziału $(0, 1)$ i statystyka (1.32) przyjmuje wartości z przedziału $(0, +\infty)$. Stąd (asymptotycznie) zbiór krytyczny dla H_0 jest prawostronny. Im lepiej więc model wyższy dopasowuje się do danych empirycznych w stosunku do modelu niższego, tym większa jest wartość statystyki ilorazu wiarygodności (1.32) i większa szansa, że wpadnie ona w przedział odrzuceń hipotezy zerowej H_0 , który leży w prawym ogonie wspomnianego rozkładu χ^2 [14, 15].

1.3.1 Dewiancja jako miara dobroci dopasowania. Rozkład Poissona.

Rozważmy zmienną losową Y posiadającą rozkład Poissona. Rozkład ten jest wykorzystywany do modelowania zjawisk związanych z rzadko zachodzącymi zdarzeniami, jak na przykład z liczbą rozpadających się niestabilnych jąder w czasie t . Ma on postać:

$$p(Y = y|\mu) = \frac{\mu^y e^{-\mu}}{y!}, \quad \text{oraz } y = 0, 1, \dots, \infty, \quad (1.33)$$

gdzie μ jest parametrem rozkładu. Zmienna losowa podlegająca rozkładowi Poissona może przyjąć tylko nieujemną wartość całkowitą. Rozkład ten można wyprowadzić z rozkładu dwumianowego, bądź wykorzystując rozkłady Erlanga i wykładniczy [11].

Związek wariancji z wartością oczekiwaną rozkład Poissona: Rozkład Poissona posiada pewną interesującą właściwość statystyczną, mianowicie jego wartość oczekiwana, wariancja i trzeci moment centralny są równe parametrowi rozkładu μ :

$$E(Y) = \sigma^2(Y) = \mu_3 = \mu. \quad (1.34)$$

Aby pokazać dwie pierwsze równości w (1.34) skorzystajmy bezpośrednio z definicji odpowiednich momentów, otrzymując:

$$\begin{aligned} E(Y) &= \sum_{y=0}^{\infty} y \cdot p(Y = y|\mu) = \sum_{y=0}^{\infty} y \cdot \frac{\mu^y e^{-\mu}}{y!} = e^{-\mu} \sum_{y=1}^{\infty} \frac{\mu^y}{(y-1)!} \\ &= e^{-\mu} \mu \sum_{y=1}^{\infty} \frac{\mu^{y-1}}{(y-1)!} = e^{-\mu} \mu \sum_{l=0}^{\infty} \frac{\mu^l}{l!} = e^{-\mu} \mu e^{\mu} = \mu, \end{aligned} \quad (1.35)$$

oraz, korzystając z (1.35):

$$\begin{aligned} \sigma^2(Y) &= E(Y^2) - [E(Y)]^2 = E(Y^2) - \mu^2 = \sum_{y=0}^{\infty} y^2 \cdot p(Y = y|\mu) - \mu^2 \\ &= \sum_{y=0}^{\infty} y^2 \cdot \frac{\mu^y e^{-\mu}}{y!} - \mu^2 = e^{-\mu} \sum_{y=1}^{\infty} y \frac{\mu^y}{(y-1)!} - \mu^2 = e^{-\mu} \mu \sum_{l=0}^{\infty} (l+1) \frac{\mu^l}{l!} - \mu^2 \\ &= e^{-\mu} \mu \left[\sum_{l=0}^{\infty} l \frac{\mu^l}{l!} + e^{\mu} \right] - \mu^2 = e^{-\mu} \mu [e^{\mu} \mu + e^{\mu}] - \mu^2 = (\mu^2 + \mu) - \mu^2 = \mu. \end{aligned} \quad (1.36)$$

Uwaga: *Zatem otrzymaliśmy ważną własność rozkładu Poissona, która mówi, że stosunek dyspersji σ do wartości oczekiwanej $E(Y)$ maleje pierwiastkowo wraz ze wzrostem poziomu zmiennej Y opisanej tym rozkładem:*

$$\frac{\sigma}{E(Y)} = \frac{1}{\sqrt{\mu}}. \quad (1.37)$$

Fakt ten oznacza z założenia *inne zachowanie się odchylenia standardowego* w modelu regresji Poissona niż w klasycznym modelu regresji normalnej (w którym zakładamy jednorodność wariancji zmiennej objaśnianej w różnych wariantach zmiennej objaśniającej).

Ćwiczenie: Pokazać (1.34) dla trzeciego momentu.

Przyczyna nielosowej zmiany wartości zmiennej objaśnianej: Rozważmy model regresji dla zmiennej objaśnianej Y posiadającej rozkład Poissona. Zmienne Y_n , $n = 1, 2, \dots, N$ posiadają więc również rozkład Poissona i zakładamy, że są *parami wzajemnie niezależne*. Niech X jest zmienną objaśniającą (tzw. czynnikiem) kontrolowanego eksperymentu, w którym X nie jest zmienną losową, ale jej *zmiana*, jest rozważana jako możliwa przyczyna warunkująca *nielosową zmianę wartości zmiennej* Y .

Gdy czynników X_1, X_2, \dots, X_k jest więcej, wtedy dla każdego punktu n próby podane są wszystkie ich wartości:

$$x_{1n}, x_{2n}, \dots, x_{kn}, \quad \text{gdzie } n = 1, 2, \dots, N, \quad (1.38)$$

gdzie pierwszy indeks w x_{in} , $i = 1, 2, \dots, k$, numeruje zmienną objaśniającą.

Brak możliwości eksperymentalnej separacji podstawowego kanału n : Niech $x_n = (x_{1n}, x_{2n}, \dots, x_{kn})$ oznacza zbiór wartości jednego wariantu zmiennych (X_1, X_2, \dots, X_k) , tzn. dla jednej konkretnej podgrupy n . Zwróćmy uwagę, że *indeks próby* n numeruje podgrupę, co oznacza, że w pomiarze wartości Y_n nie ma możliwości eksperymentalnego sięgnięcia “w głąb” indeksu n - tego kanału, tzn. do rozróżnienia wpływów na wartość y_n płynących z różnych “pod-kanałów” i , gdzie $i = 1, 2, \dots, k$.

Model podstawowy: Zakładając brak zależności zmiennej Y od czynników X_1, X_2, \dots, X_k , rozważa się tzw. *model podstawowy*. Dla rozkładu (1.33) i próby $\tilde{Y} \equiv (Y_n)_{n=1}^N$, funkcja wiarygodności przy parametrze $\Theta = \mu \equiv (\mu_n)_{n=1}^N$, ma postać:

$$P(\tilde{Y}|\mu) = \prod_{n=1}^N \frac{\mu_n^{Y_n} e^{-\mu_n}}{Y_n!} = \frac{\left(\prod_{n=1}^N \mu_n^{Y_n}\right) \exp\left(-\sum_{n=1}^N \mu_n\right)}{\prod_{n=1}^N Y_n!}, \quad (1.39)$$

jest więc wyrażona jako funkcja wektorowego parametru $\mu \equiv (\mu_n)_{n=1}^N$, gdzie każdy z parametrów $\mu_n = E(Y_n)$ jest parametrem skalarnym. Rozważmy układ równań MNW:

$$\frac{\partial}{\partial \mu_n} \left[\ln P(\tilde{Y}|\mu) \right] = 0, \quad n = 1, 2, \dots, N. \quad (1.40)$$

Dla funkcji wiarygodności (1.39) otrzymujemy:

$$\ln P(\tilde{Y}|\mu) = \sum_{n=1}^N Y_n \ln \mu_n - \sum_{n=1}^N \mu_n - \sum_{n=1}^N \ln Y_n!. \quad (1.41)$$

Zatem rozwiązanie układu (1.40) daje:

$$\mu_n = \hat{\mu}_n = Y_n, \quad n = 1, 2, \dots, N, \quad (1.42)$$

jako estymatory modelu podstawowego. Zatem funkcja wiarygodności (1.39) modelu podstawowego przyjmuje w punkcie μ zadanym przez estymatory (1.42) wartość maksymalną:

$$P(\tilde{Y}|\hat{\mu}) = \frac{\left(\prod_{n=1}^N Y_n^{Y_n}\right) \exp\left(-\sum_{n=1}^N Y_n\right)}{\prod_{n=1}^N Y_n!}, \quad (1.43)$$

gdzie zastosowano oznaczenie $\hat{\mu} = (\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_N)$.

1.3.2 Analiza regresji Poissona.

Niech zmienna zależna Y reprezentuje liczbę zliczeń badanego zjawiska (np. przypadków awarii określonego zakupionego sprzętu), otrzymaną dla każdej z N podgrup (np. klienckich). Każda z tych podgrup wyznaczona jest przez komplet wartości zmiennych objaśniających $X \equiv (X_1, X_2, \dots, X_k) = x \equiv (x_1, x_2, \dots, x_k)$ (np. wiek, poziom wykształcenia, cel nabycia sprzętu). Zmienna Y_n określa liczbę zliczeń zjawiska w n -tej podgrupie, $n = 1, 2, \dots, N$. W konkretnej próbie $(Y_n)_{n=1}^N = (y_n)_{n=1}^N$.

Określenie modelu regresji Poissona: Rozważmy następujący model regresji Poissona:

$$\mu_n \equiv E(Y_n) = \ell_n r(x_n, \beta), \quad n = 1, 2, \dots, N, \quad (1.44)$$

opisujący zmianę wartości oczekiwanej liczby zdarzeń Y_n (dla rozkładu Poissona) wraz ze zmianą *variantu* $x_n = (x_{1n}, x_{2n}, \dots, x_{kn})$.

Funkcja regresji po prawej stronie (1.44) ma dwa czynniki. Funkcyjny czynnik funkcji regresji, $r(x_n, \beta)$, opisuje *tempo zdarzeń* określanych mianem porażek (np. awarii) w n -tej podgrupie (tzn. jest częstością tego zjawiska), skąd $r(x_n, \beta) > 0$, gdzie $\beta \equiv (\beta_0, \beta_1, \dots, \beta_k)$ jest zbiorem nieznanymi parametrów tego modelu regresji. Natomiast czynnik ℓ_n jest współczynnikiem określającym *dla każdej n-tej podgrupy* (np. klientów) *skumulowany czas prowadzenia badań kontrolnych dla wszystkich jednostek tej podgrupy*.

Ponieważ funkcja regresji⁶ $r(x_n, \beta)$ przedstawia typową liczbę porażek na jednostkę czasu, zatem nazywamy ją *ryzykiem*.

Uwaga o postaci funkcji regresji: Funkcję $r(x_n, \beta)$ można zamodelować na różne sposoby [13]. Wprowadźmy oznaczenie:

$$\lambda_n^* \equiv \beta_0 + \sum_{j=1}^k \beta_j x_{jn}. \quad (1.45)$$

Funkcja regresji $r(x_n, \beta)$ ma różną postać w zależności od typu danych. Może mieć ona postać charakterystyczną dla regresji liniowej (wielokrotnej), $r(x_n, \beta) = \lambda_n^*$, którą stosujemy szczególnie wtedy gdy zmienna Y ma *rozkład normalny*. Postać $r(x_n, \beta) = 1/\lambda_n^*$ jest stosowana w analizie z danymi pochodzącymi z *rozkładu eksponentialnego*, natomiast $r(x_n, \beta) = 1/(1 + \exp(-\lambda_n^*))$ w modelowaniu regresji logistycznej dla opisu zmiennej *dychotomicznej* [14, 13].

Postać funkcji regresji użyteczna w regresji Poissona jest następująca:

$$r(x_n, \beta) = \exp(\lambda_n^*), \quad \lambda_n^* = \beta_0 + \sum_{j=1}^k \beta_j x_{jn}. \quad (1.46)$$

Ogólniej mówiąc analiza regresji odnosi się do modelowania wartości oczekiwanej zmiennej zależnej (objaśnianej) jako funkcji pewnych czynników. Postać funkcji wiarygodności stosowanej do estymacji współczynników regresji β odpowiada założeniom dotyczącym rozkładu zmiennej zależnej. Tzn. zastosowanie konkretnej funkcji regresji $r(x_n, \beta)$, np. jak w (1.46), wymaga określenia postaci funkcji

⁶Czynnik $r(x_n, \beta)$ nazywany dalej funkcją regresji, chociaż właściwie nazwa ta odnosi się do całej $E(Y_n)$.

częstości $r(x_n, \beta)$, zgodnie z jej postacią dobraną do charakteru losowej zmiennej Y przy której generowane są dane w badanym zjawisku. Na ogół przy konstrukcji $r(x_n, \beta)$ pomocna jest uprzednia wiedza dotycząca relacji między rozważanymi zmiennymi.

Funkcja wiarygodności dla analizy regresji Poissona: Ponieważ Y_n ma rozkład Poissona (1.33) ze średnią μ_n , $p(Y_n|\mu_n) = \frac{\mu_n^{Y_n}}{Y_n!} e^{-\mu_n}$, $n = 1, 2, \dots, N$, zatem dane $Y_n = 0, 1, \dots, \infty$ dla określonego $n = 1, 2, \dots, N$ są generowane z rozkładów warunkowych:

$$p(Y_n|\beta) = \frac{[\ell_n r(x_n, \beta)]^{Y_n}}{Y_n!} e^{-\ell_n r(x_n, \beta)}, \quad (1.47)$$

wokół funkcji regresji, (1.44), $\mu_n = \ell_n r(x_n, \beta)$, dla $n = 1, 2, \dots, N$. Funkcja wiarygodności dla analizy regresji Poissona ma więc postać:

$$\begin{aligned} P(\tilde{Y}|\beta) &= \prod_{n=1}^N p(Y_n|\beta) = \prod_{n=1}^N \frac{(\ell_n r(x_n, \beta))^{Y_n} e^{-\ell_n r(x_n, \beta)}}{Y_n!} \\ &= \frac{\prod_{n=1}^N (\ell_n r(x_n, \beta))^{Y_n} \exp\left[-\sum_{n=1}^N \ell_n r(x_n, \beta)\right]}{\prod_{n=1}^N Y_n!}. \end{aligned} \quad (1.48)$$

Aby w praktyce posłużyć się funkcją regresji $r(x_n, \beta)$ będącą określoną funkcją zmiennej $\lambda_n^* = \beta_0 + \sum_{j=1}^k \beta_j x_{jn}$, parametry $\beta_0, \beta_1, \dots, \beta_k$ muszą być oszacowane. Estymatory MNW, $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$, tych parametrów otrzymuje się rozwiązując $k+1$ równań wiarygodności:

$$\frac{\partial}{\partial \beta_j} \ln P(\tilde{Y}|\beta) = 0, \quad j = 0, 1, 2, \dots, k. \quad (1.49)$$

W przypadku regresji Poissona $P(\tilde{Y}|\beta)$ jest określona zgodnie z (1.48).

Algorytmy IRLS: Zauważmy, że dla rozkładu Poissona zachodzi zgodnie z (1.34) oraz (1.44), $\sigma^2(Y_n) = E(Y_n) = \ell_n r(x_n, \beta)$, co oznacza, że wariancja $\sigma^2(Y_n)$ zmiennej objaśnianej nie jest stała lecz zmienia się jako funkcja ℓ_n oraz x_n , wchodząc w analizę z różnymi wagami wraz ze zmianą n . Na fakt ten zwróciliśmy już uwagę przy okazji związku (1.37). Ponieważ układ równań wiarygodności (1.49) jest na ogół rozwiązywany iteracyjnymi metodami numerycznymi [14], a wariancja $\sigma^2(Y_n)$ jest również funkcją β , zatem na każdym kroku procesu iteracyjnego wagi te zmieniają się jako funkcja zmieniających się składowych estymatorów $\hat{\beta}$. Algorytmy takiej analizy określa się ogólnym mianem *algorytmów najmniejszych kwadratów⁷ iteracyjnie ważonych* (IRLS).

Uwaga o programach: Różne programy do analiz statystycznych, w tym SAS wykorzystujący procedurę PROC GENMOD, mogą być użyte do znajdowania estymatorów $\hat{\beta}$ MNW dla funkcji wiarygodności (1.48). Również obserwowana macierz kowariancji estymatorów⁸ oraz miary dobroci do-

⁷Należy jednak pamiętać, że zwrotu “najmniejszych kwadratów” nie należy tu brać dosłownie, gdyż metoda najmniejszych kwadratów ma sens jedynie wtedy, gdy rozkład zmiennej Y jest normalny (por. Rozdział 1.3).

⁸Obserwowana macierz (wariancji-) kowariancji $\hat{V}(\hat{\beta})$ estymatorów $\hat{\beta}$ MNW jest zdefiniowana jako odwrotność macierzy obserwowanej informacji Fishera (2.4) [13]:

$$\hat{V}(\hat{\beta}) := \mathbf{IF}^{-1}(\hat{\beta}). \quad (1.50)$$

pasowania modelu, takie jak omówiona dalej dewiancja, mogą być otrzymane przy użyciu powyżej wspomnianych programów.

1.3.2.1 Test statystyczny dla doboru modelu w regresji Poissona

Uwaga o większej wiarygodności modelu podstawowego: Maksymalna wartość funkcji wiarygodności $P(y|\mu)$ wyznaczona w oparciu o (1.43) będzie, dla każdego zbioru danych i dla liczby parametrów $k + 1 < N$, większa niż otrzymana przez maksymalizację funkcji wiarygodności (1.48). Jest tak, ponieważ w wyrażeniu (1.43) na funkcję wiarygodności modelu podstawowego *nie narzuca się żadnych ograniczeń na postać* μ_n , natomiast (1.48) wymaga aby $\mu_n = \ell_n r(x_n, \beta)$.

Hipoteza zerowa o nie występowaniu braku dopasowania w modelu niższym: Zgodnie z powyższym zdaniem, analizę doboru modelu regresji można rozpocząć od postawienia hipotezy zerowej wobec alternatywnej. W hipotezie zerowej wyróżnimy proponowany model regresji. Wybór modelu badanego oznacza wybór funkcji wiarygodności (1.48) z nim związanej. Stawiamy więc hipotezę zerową:

$$H_0 : \mu_n = \ell_n r(x_n, \beta), n = 1, 2, \dots, N, \quad (1.51)$$

która odpowiada wyborowi modelu z funkcją wiarygodności (1.48), wobec hipotezy alternatywnej:

$$H_A : \mu_n \text{ nie ma ograniczonej postaci, } n = 1, 2, \dots, N, \quad (1.52)$$

która odpowiada wyborowi modelu podstawowego zawierającego tyle parametrów μ_n ile jest punktów pomiarowych, tzn. N , z funkcją wiarygodności (1.43).

Niech więc $P(\tilde{Y}|\hat{\beta})$ jest maksymalną wartością funkcji wiarygodności określoną jak w (1.48). Oznacza to, że w miejsce parametrów $\beta = (\beta_0, \beta_1, \dots, \beta_k)$ podstawiono ich estymatory $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k)$ wyznaczone przez MNW, jako te które maksymalizują funkcję wiarygodności (1.48). Podobnie rozumiemy funkcję wiarygodności $P(\tilde{Y}|\hat{\mu})$ modelu podstawowego.

Ponieważ celem każdej analizy jest otrzymanie możliwie najprostszego opisu danych, model $\mu_n = \ell_n r(x_n, \beta)$ zawierający $k + 1$ parametrów β , będzie uznany za dobry, jeśli maksymalna wartość funkcji wiarygodności wyznaczona dla niego, będzie prawie tak duża, jak funkcji wiarygodności dla nie niosącego żadnej informacji modelu podstawowego z liczbą parametrów μ_n równą licznie punktów pomiarowych N . Sformułowanie „prawie tak duża” oznacza, że wartość funkcji wiarygodności $P(y|\hat{\beta})$ nie może być istotnie statystycznie mniejsza od $P(y|\hat{\mu})$. Zasadniczo powinno to oznaczać, że musimy podać miary pozwalające na określenie statystycznej istotności przy posługiwaniu się intuicyjnym parametrem obciążenia c (Rozdział 1.2). Okazuje się, że dla dużej próby, miary typu (1.53), podane poniżej, uzyskują cechy pozwalające na budowanie wiarygodnościowych obszarów krytycznych nabywających charakteru standardowego (częstotliwościowego).

Określenie dewiancji: Wprowadźmy statystykę typu ilorazu wiarygodności:

$$D(\hat{\beta}) = -2 \ln \left[\frac{P(\tilde{Y}|\hat{\beta})}{P(\tilde{Y}|\hat{\mu})} \right] \quad (1.53)$$

nazywaną *dewiancją* (deviance) dla modelu regresji, w tym przypadku dla modelu Poissona z określoną postacią $\mu_n = \ell_{nr}(x_n, \beta)$. Służy ona do badania dobroci dopasowania modelu zadaną postacią $\mu_n = \ell_{nr}(x_n, \beta)$ w stosunku do modelu podstawowego, bez narzuconej postaci na μ_n , tzn. do stwierdzenia, czy $P(y|\hat{\beta})$ jest istotnie *mniejsza* od $P(y|\hat{\mu})$, co sugerowałoby istotny statystycznie brak dopasowania badanego modelu $\mu_n = \ell_{nr}(x_n, \beta)$, do danych empirycznych. Jak pokażemy poniżej dewiancja może być rozumiana jako *miara zmienności reszt* (tzn. odchylenia wartości obserwowanych w próbie od wartości szacowanych przez model) *wokół linii regresji*.

Przy prawdziwości hipotezy $H_0 : \mu_n = \ell_{nr}(x_n, \beta)$, rozkład dewiancji $D(\hat{\beta})$ dla regresji Poissona, można asymptotycznie przybliżyć rozkładem chi-kwadrat (por. dyskusja w [13, 14]) z $N - k - 1$ stopniami swobody.

Zatem statystyczny test dobroci dopasowania, tzn. niewystępowania braku dopasowania badanego modelu $H_0 : \mu_n = \ell_{nr}(x_n, \beta)$, w stosunku do modelu podstawowego, przebiega w regresji Poissona następująco: Porównujemy otrzymaną w próbie wartość statystyki $D(\hat{\beta})$ z wartością krytyczną leżącą w prawym ogonie rozkładu chi-kwadrat (o $N - k - 1$ stopniach swobody). Przyjęcie przez $D(\hat{\beta})$ wartości równej lub większej od krytycznej skutkuje odrzuceniem hipotezy zerowej.

Wyznaczenie liczby stopni swobody dewiancji: Podana liczba stopni swobody dewiancji $D(\hat{\beta})$ wynika z następującego rozumowania. Zapiszmy (1.53) w postaci:

$$D(\hat{\beta}) + 2 \ln P(\tilde{Y}|\hat{\beta}) = 2 \ln P(\tilde{Y}|\hat{\mu}) , \quad (1.54)$$

co po skorzystaniu z (1.48) dla $\beta = \hat{\beta}$ ma postać:

$$D(\hat{\beta}) - 2 \sum_{n=1}^N \ell_{nr}(x_n, \hat{\beta}) = 2 \ln P(\tilde{Y}|\hat{\mu}) + 2 \ln \left(\prod_{n=1}^N Y_n! \right) - 2 \ln \left(\prod_{n=1}^N \left(\ell_{nr}(x_n, \hat{\beta}) \right)^{Y_n} \right) . \quad (1.55)$$

Można zauważyć, że prawa strona tego równania ma N -stopni swobody. Istotnie, ze względu na (1.42)⁹, $\hat{\mu} \equiv (\hat{\mu}_n) = (Y_n)$, $n = 1, 2, \dots, N$, liczba niezależnych zmiennych po prawej strony powyższego równania, których wartości trzeba określić z eksperymentu, wynosi N . Natomiast drugi składnik po lewej stronie ma liczbę stopni swobody równą $k + 1$, co jest liczbą estymatorów parametrów strukturalnych $\hat{\beta}$ modelu regresji, których wartości trzeba określić z eksperymentu. Ponieważ liczba stopni swobody po prawej i lewej stronie równania musi być taka sama, zatem liczba stopni swobody dewiancji $D(\hat{\beta})$ wynosi $N - k - 1$.

Testy ilorazu wiarygodności: Dewiancje dla hierarchicznych klas modeli mogą służyć do budowy testów stosunku wiarygodności. Zwróćmy szczególnie uwagę na funkcję wiarygodności (1.48) zawierającą zbiór parametrów $\beta = (\beta_0, \beta_1, \dots, \beta_k)$ z *dewiancją* $D(\hat{\beta})$ daną wyrażeniem (1.53). Przypuśćmy, że chcemy zweryfikować hipotezę o tym, że $k - r$ (gdzie $0 < r < k$) ostatnich parametrów będących składowymi wektora β jest równych zeru.

Hipoteza zerowa, o nieistotności rozszerzenia modelu niższego do wyższego, ma wtedy postać:

$$H_0 : \beta_{r+1} = \beta_{r+2} = \dots = \beta_k = 0 , \quad (1.56)$$

⁹W przyjętym przedstawieniu danych jak dla diagramu punktowego, N jest ogólnie liczbą punktów pomiarowych (równą liczbie wariantów czy komórek). Tylko dla modelu podstawowego jest N również liczbą parametrów.

Hipoteza alternatywna H_A mówi, że przynajmniej jeden z parametrów strukturalnych $\beta_{r+1}, \beta_{r+2}, \dots, \beta_k$ jest różny od zera.

Funkcja wiarygodności przy prawdziwości hipotezy zerowej H_0 , (1.56), ma postać taką jak w (1.48), tyle, że zastąpiono w niej parametr β parametrem $\beta_{(r)}$:

$$\beta_{(r)} \equiv (\beta_0, \beta_1, \dots, \beta_r; 0, 0, \dots, 0) \text{ gdzie liczba zer wynosi } k - r. \quad (1.57)$$

Oznaczmy funkcję wiarygodności tego modelu jako $P(\tilde{Y}|\beta_{(r)})$, a $\hat{\beta}_{(r)}$ niech będzie estymatorem MNW wektorowego parametru $\beta_{(r)}$, wyznaczonym przez rozwiązanie odpowiadającego mu układu równań wiarygodności (oczywiście dla niezerowych parametrów $\beta_0, \beta_1, \dots, \beta_r$). Estymator $\hat{\beta}_{(r)} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_r; 0, 0, \dots, 0)$ maksymalizuje funkcję wiarygodności $P(\tilde{Y}|\beta_{(r)})$.

Test ilorazu wiarygodności dla weryfikacji hipotezy H_0 przeprowadzamy posługując się *statystyką ilorazu wiarygodności*:

$$-2 \ln \left[\frac{P(\tilde{Y}|\hat{\beta}_{(r)})}{P(\tilde{Y}|\hat{\beta})} \right], \quad (1.58)$$

która przy prawdziwości hipotezy zerowej ma asymptotycznie rozkład chi-kwadrat z $k - r$ stopniami swobody, co widać, gdy zapiszemy (1.58) jako różnicę dewiancji:

$$-2 \ln \left[\frac{P(\tilde{Y}|\hat{\beta}_{(r)})}{P(\tilde{Y}|\hat{\beta})} \right] = -2 \ln \left[\frac{P(\tilde{Y}|\hat{\beta}_{(r)})}{P(\tilde{Y}|\hat{\mu})} \right] + 2 \ln \left[\frac{P(\tilde{Y}|\hat{\beta})}{P(\tilde{Y}|\hat{\mu})} \right] = D(\hat{\beta}_{(r)}) - D(\hat{\beta}), \quad (1.59)$$

oraz skorzystamy z podobnej analizy jak dla (1.55).

Zatem, przy prawdziwości hipotezy zerowej (1.56), którą można zapisać jako $H_0 : \beta_{r+1} = \beta_{r+2} = \dots = \beta_k = 0$, różnica $D(\hat{\beta}_{(r)}) - D(\hat{\beta})$ ma dla dużej próby w przybliżeniu rozkład chi-kwadrat z $k - r$ stopniami swobody.

Wniosek: Jeśli używamy regresji Poissona do analizowania danych empirycznych, modele tworzące hierarchiczne klasy mogą być porównywane między sobą poprzez wyznaczenie statystyki ilorazu wiarygodności (1.58), lub co na jedno wychodzi, poprzez wyznaczenie różnicy (1.59) między parami dewiancji dla tych modeli. Należy przy tym pamiętać o wniosku jaki już znamy z analizy dewiancji, że *im model gorzej dopasowuje się do danych empirycznych tym jego dewiancja jest większa*.

1.3.2.2 Podobieństwo dewiancji do SKR analizy częstotliwościowej

Warunkowe wartości oczekiwane $\mu_n \equiv E(Y_n) = \ell_n r(x_n, \beta)$, $n = 1, 2, \dots, N$, (1.44), są w analizie regresji przyjmowane jako teoretyczne przewidywania modelu regresji dla wartości zmiennej objaśnianej Y_n , zwanej odpowiedzią (układu). W próbie odpowiadają im oszacowania, oznaczone jako \hat{Y}_n , które w n -tej komórce są następujące:

$$\hat{Y}_n = \ell_n r(x_n, \hat{\beta}), \quad n = 1, 2, \dots, N, \quad (1.60)$$

zgodnie z wyestymowaną postacią modelu regresji. Wykorzystując (1.60) możemy zapisać dewiancję modelu (1.53) następująco:

$$\begin{aligned} D(\hat{\beta}) &= -2 \ln \left[\frac{P(\tilde{Y}|\hat{\beta})}{P(\tilde{Y}|\hat{\mu})} \right] = -2 \ln \left[\frac{\prod_{n=1}^N \hat{Y}_n^{Y_n} \exp\left(-\sum_{n=1}^N \hat{Y}_n\right)}{\prod_{n=1}^N Y_n^{Y_n} \exp\left(-\sum_{n=1}^N Y_n\right)} \right] \\ &= -2 \left[\sum_{n=1}^N Y_n \ln \hat{Y}_n - \sum_{n=1}^N \hat{Y}_n - \sum_{n=1}^N Y_n \ln Y_n + \sum_{n=1}^N Y_n \right] \end{aligned} \quad (1.61)$$

tzn:

$$D(\hat{\beta}) = 2 \sum_{n=1}^N \left[Y_n \ln \left(\frac{Y_n}{\hat{Y}_n} \right) - (Y_n - \hat{Y}_n) \right]. \quad (1.62)$$

Podobieństwo D do SKR : Powyższa postać dewiancji oznacza, że $D(\hat{\beta})$ zachowuje się w poniższym sensie jak suma kwadratów reszt $SKR = \sum_{n=1}^N (Y_n - \hat{Y}_n)^2$ w standardowej wielorakiej regresji liniowej. Otóż, gdy dopasowywany model dokładnie przewiduje obserwowane wartości, tzn. $\hat{Y}_n = Y_n$, $n = 1, 2, \dots, N$ wtedy, jak SKR w analizie standardowej, tak $D(\hat{\beta})$ w analizie wiarygodnościowej jest równe zeru [14, 17]. Z drugiej strony wartość $D(\hat{\beta})$ jest tym większa im większa jest różnica między wartościami obserwowanymi Y_n i wartościami przewidywanymi \hat{Y}_n przez oszacowany model.

Asymptotyczna postać D : W analizowanym modelu Y_n , $n = 1, 2, \dots, N$ są niezależnymi zmiennymi Poissona (np. zmiennymi częstości), natomiast wartości \hat{Y}_n są ich przewidywaniami. Nietrudno przekonać się, że gdy wartości przewidywane mają rozsądną wartość¹⁰, np. $\hat{Y}_n > 3$ oraz $(Y_n - \hat{Y}_n) \ll Y_n$, $n = 1, 2, \dots, N$ tak, że $(Y_n - \hat{Y}_n)/Y_n \ll 1$, wtedy wyrażenie w nawiasie kwadratowym w (1.62) można przybliżyć przez $(Y_n - \hat{Y}_n)^2/(2 Y_n)$, a statystykę (1.62) można przybliżyć statystyką o postaci:

$$\chi^2 = \sum_{n=1}^N \frac{(Y_n - \hat{Y}_n)^2}{\hat{Y}_n}, \quad (1.63)$$

która (dla dużej próby) ma rozkład chi-kwadrat z $N - k - 1$ stopniami swobody [14].

¹⁰Zauważmy, że statystyka (1.63) może mieć myląco dużą wartość gdy wielkości \hat{Y}_n są bardzo małe.

1.4 Zasada niezmienniczości ilorazu funkcji wiarygodności

Z powyższych rozważań wynika, że funkcja wiarygodności reprezentuje niepewność dla ustalonego parametru. Nie jest ona jednak gęstością rozkładu prawdopodobieństwa dla tego parametru. Pojęcie takie byłoby całkowicie obce statystyce klasycznej (nie włączając procesów stochastycznych). Inaczej ma się sprawa w tzw. statystyce Bayesowskiej. Aby zrozumieć różnicę pomiędzy podejściem klasycznym i Bayesowskim [18] rozważmy transformację parametru.

Przykład transformacji parametru: Rozważmy eksperyment, w którym dokonujemy jednokrotnego pomiaru zmiennej o rozkładzie dwumianowym (1.10). Funkcja wiarygodności ma więc postać $P(\theta) = \binom{m}{x} \theta^x (1 - \theta)^{m-x}$. Niech parametr $m = 12$ a w pomiarze otrzymano $x = 9$. Testujemy model, dla którego $\theta = \theta_1 = 3/4$ wobec modelu z $\theta = \theta_2 = 3/10$. Stosunek wiarygodności wynosi:

$$\frac{P(\theta_1 = 3/4)}{P(\theta_2 = 3/10)} = \frac{\binom{m}{x} \theta_1^9 (1 - \theta_1)^3}{\binom{m}{x} \theta_2^9 (1 - \theta_2)^3} = 173.774 \quad (1.64)$$

Dokonajmy hiperbolicznego wzajemnie jednoznacznego przekształcenia parametru:

$$\psi = 1/\theta. \quad (1.65)$$

Funkcja wiarygodności po transformacji parametru ma postać $\tilde{P}(\psi) = \binom{m}{x} (1/\psi)^x (1 - 1/\psi)^{m-x}$.

Wartości parametru ψ odpowiadające wartościom θ_1 i θ_2 wynoszą odpowiednio $\psi_1 = 4/3$ oraz $\psi_2 = 10/3$. Łatwo sprawdzić, że transformacja (1.65) nie zmienia stosunku wiarygodności, tzn.:

$$\frac{\tilde{P}(\psi_1 = 4/3)}{\tilde{P}(\psi_2 = 10/3)} = \frac{P(\theta_1 = 3/4)}{P(\theta_2 = 3/10)} = 173.774. \quad (1.66)$$

Niezmienniczość stosunku wiarygodności: Zatem widać, że stosunek wiarygodności jest niezmienny ze względu na wzajemnie jednoznaczną transformację parametru. Gdyby transformacja parametru była np. transformacją "logit" $\psi = \ln(\theta/(1 - \theta))$ lub paraboliczną $\psi = \theta^2$, to sytuacja także nie uległaby zmianie. Również w ogólnym przypadku transformacji parametru własność *niezmienniczości stosunku wiarygodności* pozostaje słuszna. Oznacza to, że informacja zawarta w próbce jest niezmiennicza ze względu na wybór parametryzacji, tzn. powinniśmy być w takiej samej sytuacji niewiedzy niezależnie od tego jak zamodelujemy zjawisko, o ile różnica w modelowaniu sprowadza się jedynie do transformacji parametru. W omawianym przykładzie powinniśmy równie dobrze móc stosować parametr θ , jak $1/\theta$, θ^2 , czy $\ln(\theta/(1 - \theta))$.

Uwaga o transformacji parametru w statystyce Bayesowskiej: Natomiast sytuacja ma się zupełnie inaczej w przypadku Bayesowskiego podejścia do funkcji wiarygodności [18], w którym funkcja wiarygodności uwzględnia (Bayesowski) rozkład prawdopodobieństwa $f(\theta|x)$ parametru θ . Oznacza to, że Jakobian transformacji $\theta \rightarrow \psi$ parametru, modyfikując rozkład parametru, zmienia również funkcję wiarygodności. Zmiana ta zależy od wartości parametru, różnie zmieniając licznik i mianownik w (1.64), co niszczy *intuicyjną* własność niezmienniczości ilorazu wiarygodności ze względu na transformację parametru [13].

Rozdział 2

Entropia względna i informacja Fishera

W pozostałych częściach skryptu nie będziemy zajmowali się modelami regresyjnymi. Oszacowywany wektorowy parametr Θ , od którego zależy funkcja wiarygodności $P(\Theta)$ ma postać $\Theta = (\theta_1, \theta_2, \dots, \theta_N)^T \equiv (\theta_n)_{n=1}^N$ jak w (1.4), gdzie $\theta_n = (\vartheta_{1n}, \vartheta_{2n}, \dots, \vartheta_{kn})^T \equiv ((\vartheta_s)_{s=1}^k)_n$, skąd liczba wszystkich parametrów wynosi $d = k \times N$. Aby nie komplikować zapisu będziemy stosowali oznaczenie $\Theta \equiv (\theta_1, \theta_2, \dots, \theta_d)^T \equiv (\theta_i)_{i=1}^d$, gdzie indeks $i = 1, 2, \dots, d$ zastąpił parę indeksów sn .

Niech $\hat{\Theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_d)$ jest estymatorem MNW wektorowego parametru $\Theta = (\theta_1, \theta_2, \dots, \theta_d)$, otrzymanym po rozwiązaniu układu równań wiarygodności (1.8). Rozwiązanie to, jako maksymalizujące funkcję wiarygodności, musi spełniać warunek ujemnej określoności formy kwadratowej¹:

$$\sum_{i,j=1}^d \frac{\partial^2 \ln P}{\partial \theta_i \partial \theta_j} \Big|_{\Theta=\hat{\Theta}} \Delta \theta_i \Delta \theta_j, \quad (2.2)$$

gdzie przyrosty $\Delta \theta_i, \Delta \theta_j$ nie zerują się jednocześnie. W przypadku skalarnym (tzn. jednego parametru $\Theta = \theta = \vartheta$) warunek ten oznacza ujemność drugiej pochodnej logarytmu funkcji wiarygodności w punkcie $\theta = \hat{\theta}$. Większa wartość $-\frac{\partial^2}{\partial \theta^2} \ln P(y|\theta) \Big|_{\theta=\hat{\theta}}$ oznacza węższe maksimum $\ln P$ w punkcie $\theta = \hat{\theta}$, tzn. większą krzywiznę funkcji wiarygodności, a co za tym idzie mniejszą niepewność określenia parametru θ .

2.1 Obserwowana i oczekiwana informacja Fishera

Ponieważ Θ może być w ogólności parametrem wektorowym, więc jako uogólnienie przypadku skalarnego zdefiniujemy $d \times d$ -wymiarową macierz²:

$$\mathbf{IF}(\Theta) \equiv -\frac{\partial^2}{\partial \Theta \partial \Theta^T} \ln P(y|\Theta) \equiv -\left(\frac{\partial^2 \ln P}{\partial \theta_i \partial \theta_j} \right)_{d \times d}. \quad (2.3)$$

¹W pełnym zapisie indeksów, ze względu na to, że jeden punktowy parametr $\theta_n = (\vartheta_{1n}, \vartheta_{2n}, \dots, \vartheta_{kn})^T \equiv ((\vartheta_s)_{s=1}^k)_n$ może być parametrem wektorowym, co ma miejsce gdy $k > 1$, zapis (2.2) oznacza:

$$\sum_{n,n'=1}^N \sum_{s,s'=1}^k \frac{\partial^2 \ln P}{\partial \vartheta_{sn} \partial \vartheta_{s'n'}} \Big|_{\Theta=\hat{\Theta}} \Delta \vartheta_{sn} \Delta \vartheta_{s'n'}. \quad (2.1)$$

Wkrótce i tak ograniczymy się do sytuacji gdy $k = 1$, tzn. $\theta_n = \vartheta_{1n} = \vartheta_n$.

² Będziemy powszechnie stosowali skrócony zapis typu: $\Theta^2 \equiv \Theta \Theta^T = (\theta_i \theta_j)_{d \times d}$ oraz $\frac{\partial^2}{\partial \Theta^2} \equiv \frac{\partial^2}{\partial \Theta \partial \Theta^T} = \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} \right)_{d \times d}$.

Określenie obserwowanej informacji Fishera: Wartość statystyki $iF(\Theta)$ zdefiniowanej jak w (2.3) w punkcie $\Theta = \hat{\Theta}$ (oznaczoną po prostu jako iF):

$$iF = iF(\hat{\Theta}) = iF(\Theta)|_{\Theta=\hat{\Theta}} , \quad (2.4)$$

nazywamy *obserwowaną informacją Fishera*. W teorii wiarygodności odgrywa ona kluczową rolę. Jako statystyka, czyli funkcja próby \tilde{Y} , jej realizacja w próbce $y \equiv (y_1, y_2, \dots, y_N)$ jest macierzą liczbową. Z faktu wyznaczenia iF w punkcie estymatora MNW wynika, że jest ona dodatnio określona, natomiast z (2.3) widać również, że iF jest macierzą symetryczną.

We współczesnych wykładach iF jest zazwyczaj zapisane jako:

$$\tilde{iF} = \left(\frac{\partial \ln P(\Theta)}{\partial \theta_{i'}} \frac{\partial \ln P(\Theta)}{\partial \theta_i} \right) . \quad (2.5)$$

Obie definicje, tzn. (2.5) oraz (2.3), prowadzą na poziomie oczekiwanym do tych samych konkluzji, o ile $\int dy P(\Theta) \frac{\partial \ln P(\Theta)}{\partial \theta_i} = 0$, $i = 1, 2, \dots, d$ (por. (2.31)). Zasadniczą zaletą zdefiniowania iF poprzez (2.5) jest to, że bardzo naturalne staje się wtedy wprowadzenie tzw. α -koneksji na przestrzeni statystycznej \mathcal{S} [6]. Pojęcie α -koneksji omówimy w Rozdziale 2.2.

Przykład estymacji wartości oczekiwanej w rozkładzie normalnym: Jako przykład ilustrujący związek wielkości obserwowanej informacji Fishera (IF) z niepewnością oszacowania parametru, rozważmy realizację próby prostej y dla zmiennej Y posiadającej rozkład $N(\theta, \sigma^2)$.

Założmy, że *wariancja σ^2 jest znana*, a *estymowanym parametrem jest jedynie wartość oczekiwana $\theta = E(Y)$* . Logarytm funkcji wiarygodności ma postać:

$$\ln P(y|\theta) = -\frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=1}^N (y_n - \theta)^2 \quad (2.6)$$

skąd funkcja wynikowa (1.9) jest równa:

$$S(\theta) = \frac{\partial}{\partial \theta} \ln P(y|\theta) = \frac{1}{\sigma^2} \sum_{n=1}^N (y_n - \theta) . \quad (2.7)$$

Rozwiązując jedno równanie wiarygodności, $S(\theta)|_{\theta=\hat{\theta}} = 0$, otrzymujemy postać estymatora parametru θ (por. (1.19)):

$$\hat{\theta} = \bar{y} = \frac{1}{N} \sum_{n=1}^N y_n , \quad (2.8)$$

skąd:

$$S(\theta) = \frac{N}{\sigma^2} (\hat{\theta} - \theta) . \quad (2.9)$$

Natomiast z (2.4) oraz (2.3) otrzymujemy, że obserwowana IF jest równa:

$$iF(\hat{\theta}) = -\frac{\partial^2 \ln P(y|\theta)}{\partial \theta^2} |_{\theta=\hat{\theta}} = -\frac{\partial S(\theta)}{\partial \theta} |_{\theta=\hat{\theta}} = \frac{N}{\sigma^2} . \quad (2.10)$$

Z (2.8) oraz z klasycznej analizy³ [11] wiemy, że wariancja:

$$\sigma^2(\hat{\theta}) = \sigma^2(\bar{y}) = \sigma^2/N , \quad (2.11)$$

zatem:

$$\text{iF}(\hat{\theta}) = \frac{1}{\sigma^2(\hat{\theta})} . \quad (2.12)$$

Wniosek: Otrzymaliśmy więc ważny związek mówiący, że większa obserwowana IF parametru θ oznacza mniejszą wariancję jego estymatora $\hat{\theta}$.

Równanie (2.12) można zapisać w postaci:

$$\sigma^2(\hat{\theta}) \text{iF}(\hat{\theta}) = 1 , \quad (2.13)$$

co w przypadku estymacji wartości oczekiwanej rozkładu normalnego, jest sygnałem osiągnięcia *dolnego ograniczenia nierówności Rao-Craméra* [6]. Temat ten będziemy rozwijać dalej.

Określenie oczekiwanej IF: Zdefiniujmy oczekiwaną informację Fishera następująco:

$$I_F(\Theta) \equiv E_{\Theta}(\text{iF}(\Theta)) = \int_{\mathcal{B}} dy P(y|\Theta) \text{iF}(\Theta) , \quad (2.14)$$

gdzie \mathcal{B} jest przestrzenią próby (układu). Oznaczenie Θ w indeksie wartości oczekiwanej mówi, że Θ jest prawdziwą wartością parametru, przy której generowane są dane $y \equiv (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)$, natomiast element różniczkowy dy oznacza:

$$dy \equiv d^N \mathbf{y} = d\mathbf{y}_1 d\mathbf{y}_2 \dots d\mathbf{y}_N . \quad (2.15)$$

Oczekiwana v.s. obserwowana IF: Istnieją znaczące różnice pomiędzy oczekiwaną a obserwowaną informacją Fishera [13, 19]. Oczekiwana informacja Fishera I_F ma sens jako funkcja dopuszczalnych wartości Θ , należących do przestrzeni V_{Θ} wartości Θ . Natomiast jak to wynika z MNW, obserwowana informacja Fishera, $\text{iF}(\Theta)$, ma zasadniczo sens tylko w pobliżu $\hat{\Theta}$. Jako związana z obserwowaną wartością wiarygodności, iF odnosi się do pojedynczego zestawu danych i zmienia się od próbki do próbki. Oznacza to, że należy o niej myśleć jako o pojedynczej realizacji statystyki $\text{iF}(\hat{\Theta})$ w próbce, a nie jako o funkcji parametru Θ . Natomiast oczekiwana informacja I_F jest średnią wartością dla wszystkich możliwych zestawów danych w całej przestrzeni próby \mathcal{B} , generowanych przy prawdziwej wartości parametru. Zatem $I_F(\Theta)$ jest nie tyle użytecznym wskaźnikiem informacji dla konkretnego zbioru danych, ile funkcją Θ mówiącą jak trudno jest estymować Θ , co oznacza, że parametr z większą I_F wymaga mniejszej próbki do osiągnięcia wymaganej precyzji jego oszacowania.

Kontynuując rozpoczęty powyżej Przykład rozkładu normalnego z estymacją skalarne parametru θ , otrzymujemy po skorzystaniu z (2.10) oraz unormowaniu funkcji wiarygodności:

$$\int dy P(y|\theta) = 1 , \quad (2.16)$$

wartość oczekiwaną IF dla parametru θ rozkładu $N(\theta, \sigma^2)$:

$$I_F(\theta) = \int dy P(y|\theta) \text{iF}(\theta) = \int dy P(y|\theta) \frac{N}{\sigma^2} = \frac{N}{\sigma^2} . \quad (2.17)$$

³Ponieważ dla próby prostej wszystkie Y_n , $n = 1, 2, \dots, N$, mają taki sam rozkład jak Y oraz dla $n \neq n'$ zachodzi $\text{cov}(Y_n, Y_{n'}) = 0$, zatem: $\sigma^2(\bar{Y}) = \sigma^2(\frac{1}{N} \sum_{n=1}^N Y_n) = \frac{1}{N^2} \sigma^2(\sum_{n=1}^N Y_n) = \frac{1}{N^2} \sum_{n=1}^N \sigma^2(Y_n) = \frac{1}{N^2} N \sigma^2(Y) = \frac{\sigma^2}{N}$.

Ze względu na (2.11) wynik ten oznacza, że zachodzi:

$$\sigma^2(\hat{\theta})I_F(\theta) = 1, \quad (2.18)$$

co mówi, że w estymacji wartości oczekiwanej rozkładu normalnego osiągamy *dolne ograniczenie nierówności Rao-Craméra* [6]. Nierówność ta daje najważniejsze ograniczenie statystyki informacyjnej na jakość estymacji w pojedynczym kanale informacyjnym. Sprawie nierówności Rao-Cramera poświęcimy część rozważań skryptu.

Przykład estymacji obu parametrów rozkładu normalnego: Rozważmy dwuparametrowy ($d = 2$) rozkład normalny. Niech $\mathbf{y}_1, \dots, \mathbf{y}_N$ jest realizacją próby prostej dla zmiennej Y o rozkładzie normalnym $N(\mu, \sigma)$. Ponieważ $\theta_i^n = \theta_i$ dla $n = 1, 2, \dots, N$ zatem wektor parametrów przyjmujemy jako $\Theta = ((\theta_i^n)_{i=1}^2)_{n=1}^N \equiv (\theta_i)_{i=1}^2 = (\mu, \sigma)^T$. Funkcja wiarygodności próby ma wtedy postać:

$$\ln P(y|\Theta) = -N \ln(\sqrt{2\pi} \sigma) - \frac{1}{2\sigma^2} \sum_{n=1}^N (\mathbf{y}_n - \mu)^2. \quad (2.19)$$

Funkcja wynikowa z nią związana jest równa:

$$S(\Theta) = \begin{pmatrix} \frac{\partial}{\partial \mu} \ln P(y|\Theta) \\ \frac{\partial}{\partial \sigma} \ln P(y|\Theta) \end{pmatrix} = \begin{pmatrix} \frac{N}{\sigma^2} (\bar{\mathbf{y}} - \mu) \\ -\frac{N}{\sigma} + \frac{\sum_{n=1}^N (\mathbf{y}_n - \mu)^2}{\sigma^3} \end{pmatrix}, \quad (2.20)$$

gdzie $\bar{\mathbf{y}} = \frac{1}{N} \sum_{n=1}^N \mathbf{y}_n$ jest średnią arytmetyczną w próbie. Zatem postacie estymatorów MNW, tzn. parametru wartości oczekiwanej μ oraz odchylenia standardowego σ zmiennej Y , otrzymujemy rozwiązując układ równań wiarygodności:

$$S(\Theta)|_{\Theta=\hat{\Theta}} = \begin{pmatrix} \frac{N}{\sigma^2} (\bar{\mathbf{y}} - \mu) \\ -\frac{N}{\sigma} + \frac{\sum_{n=1}^N (\mathbf{y}_n - \mu)^2}{\sigma^3} \end{pmatrix} |_{\Theta=\hat{\Theta}} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad (2.21)$$

gdzie $\hat{\Theta} = (\hat{\mu}, \hat{\sigma})^T$ i rozwiązanie to ma postać:

$$\hat{\mu} = \bar{\mathbf{y}} \quad \text{oraz} \quad \hat{\sigma} = \sqrt{\frac{1}{N} \sum_{n=1}^N (\mathbf{y}_n - \bar{\mathbf{y}})^2}. \quad (2.22)$$

Obserwowana informacja Fishera (2.3) w punkcie $\hat{\Theta}$ wynosi więc:

$$\mathbf{iF}(\hat{\Theta}) = \begin{pmatrix} \frac{N}{\sigma^2} & \frac{2N}{\sigma^3} (\bar{\mathbf{y}} - \mu) \\ \frac{2N}{\sigma^3} (\bar{\mathbf{y}} - \mu) & -\frac{N}{\sigma^2} + \frac{3}{\sigma^4} \sum_{n=1}^N (\mathbf{y}_n - \mu)^2 \end{pmatrix} |_{\Theta=\hat{\Theta}} = \begin{pmatrix} \frac{N}{\sigma^2} & 0 \\ 0 & \frac{2N}{\sigma^2} \end{pmatrix}, \quad (2.23)$$

W końcu oczekiwana informacja Fishera jest równa:

$$I_F(\Theta) = E_{\Theta} \mathbf{iF}(\Theta) = \int_{\mathcal{B}} dy P(y|\Theta) \mathbf{iF}(\Theta) = \begin{pmatrix} \frac{N}{\sigma^2} & 0 \\ 0 & \frac{2N}{\sigma^2} \end{pmatrix}. \quad (2.24)$$

Uwaga o estymatorze $\hat{\sigma}^2$ MNW: Widząc powyższy rozkład jako $N(\mu, \sigma^2)$ i w konsekwencji przyjmując wektor parametrów jako $(\mu, \sigma^2)^T$, oraz przeprowadzając analogiczny rachunek jak powyżej, można pokazać, że estymator wartości oczekiwanej ma postać $\hat{\mu} = \bar{\mathbf{y}}$, czyli tak jak w (2.22), natomiast estymator wariancji wynosi $\hat{\sigma}^2 = \frac{1}{N} \sum_{n=1}^N (\mathbf{y}_n - \bar{\mathbf{y}})^2$. W jego mianowniku występuje czynnik N a

nie $N - 1$ jak to ma miejsce w przypadku nieobciążonego estymatora wariancji. Estymator $\hat{\sigma}^2$ nie jest też efektywny (tzn. nie posiada najmniejszej z możliwych wariancji). Jednak własności te posiada on asymptotycznie ($N \rightarrow \infty$), co jest charakterystyczne dla wszystkich estymatorów MNW [11]. Sprawdzenie, że przy odpowiednio dobranych tzw. oczekiwanych parametrach rozkładu normalnego, estymacja obu parametrów jest efektywna dla skończonego N , pozostawiamy jako ćwiczenie na koniec tego rozdziału, po tym jak zapoznamy się z pojęciem dualnych układów współrzędnych.

2.1.1 Wartość oczekiwana i wariancja funkcji wynikowej

Pokażmy, że wartość oczekiwana funkcji wynikowej $S(\Theta) \equiv S(\tilde{Y}|\Theta)$, tzn. gradientu logarytmu naturalnego funkcji wiarygodności, jest równa zero:

$$E_{\Theta} S(\Theta) = 0. \quad (2.25)$$

Istotnie, gdy skorzystamy z interpretacji funkcji wiarygodności jako łącznego rozkładu prawdopodobieństwa i jej unormowania do jedności, wtedy:

$$\begin{aligned} E_{\Theta} S(\Theta) &= \int_{\mathcal{B}} dy P(y|\Theta) S(\Theta) = \int_{\mathcal{B}} dy P(y|\Theta) \left(\frac{\partial}{\partial \Theta} \ln P(y|\Theta) \right) \\ &= \int_{\mathcal{B}} dy P(y|\Theta) \frac{\frac{\partial}{\partial \Theta} P(\Theta)}{P(\Theta)} = \int_{\mathcal{B}} dy \frac{\partial}{\partial \Theta} P(y|\Theta) = \frac{\partial}{\partial \Theta} \int_{\mathcal{B}} dy P(y|\Theta) = 0, \end{aligned} \quad (2.26)$$

gdzie zakres całkowania obejmuje całą przestrzeń próby \mathcal{B} . W przypadku gdy pierwotna zmienna Y jest dyskretna, powyższy dowód przebiega podobnie.

Założenie o regularności rozkładu: W (2.26) wyciągnęliśmy różniczkowanie po parametrze przed znak całki. Poprawność takiego przejścia oznacza spełnienie żądania, aby rozkład $P(y|\Theta)$ był wystarczająco gładki jako funkcja Θ [13]. Oznacza to, że Θ nie może być brzegową wartością, tzn. istnieje taka funkcja $g(y)$ (której całka $\int_{\mathcal{B}} g(y) dy$ jest skończona), dla której w sąsiedztwie prawdziwej wartości parametru Θ zachodzi $|\partial P(y|\Theta)/\partial \Theta| \leq g(y)$ gdy $\partial P/\partial \Theta$ jest traktowana jako funkcja y [13], skąd wynika skończoność całki $\int_{\mathcal{B}} dy \partial P(y|\Theta)/\partial \Theta$.

Zauważmy również, że przy odpowiednim warunku regularności:

$$\int_{\mathcal{B}} dy \frac{\partial^2}{\partial \Theta^2} P(y|\Theta) = \frac{\partial^2}{\partial \Theta^2} \int_{\mathcal{B}} dy P(y|\Theta) = 0. \quad (2.27)$$

Twierdzenie o wariancji funkcji wynikowej: Zakładając warunek regularności pozwalający na wyciągnięcie różniczkowania po parametrze przed znak całki, można pokazać, że wariancja (precyzyjnie, macierz wariancji-kowariancji) funkcji wynikowej $S(\Theta) \equiv S(\tilde{Y}|\Theta)$ jest równa oczekiwanej IF⁴:

$$\sigma_{\Theta}^2 S(\Theta) = I_F(\Theta). \quad (2.28)$$

Zauważmy, że z powyższego wynika, że $d \times d$ -wymiarowa macierz informacji $I_F(\Theta)$ jest macierzą kowariancji, co oznacza, że jest ona *nieujemnie określona*⁵.

⁴Ponieważ funkcja wynikowa $S(\Theta) \equiv S(\tilde{Y}|\Theta) = \frac{\partial P(\tilde{Y}|\Theta)}{\partial \Theta}$ jest d -wymiarowym kolumnowym wektorem losowym (1.9) z wartością oczekiwaną równą zero, zatem $\text{Cov}_{\Theta} \left[S(\tilde{Y}|\Theta) \right]$ jest $d \times d$ -wymiarową macierzą kowariancji (współrzędnych) wektora $S(\tilde{Y}|\Theta)$. W skrypcie będziemy stosowali oznaczenie $\sigma_{\Theta}^2 S(\Theta) \equiv \text{Cov}_{\Theta} \left[S(\tilde{Y}|\Theta) \right]$.

⁵Dla dowolnego wektora $\mathbf{a} = (a_1, \dots, a_d)^T \in \mathbf{R}^d$ oraz macierzy kowariancji $C = E \left((Z - E(Z))(Z - E(Z))^T \right)$, gdzie Z jest d -wymiarowym wektorem losowym, zachodzi $\mathbf{a}^T E \left[(Z - E(Z))(Z - E(Z))^T \right] \mathbf{a} = \mathbf{a}^T E(WW^T) \mathbf{a} = E[(\mathbf{a}^T W)(\mathbf{a}^T W)^T] = \sigma^2(\mathbf{a}^T W) \geq 0$, gdzie $W = Z - E(Z)$ i $E(W) = 0$, tzn. macierz C jest nieujemnie określona.

Dowód twierdzenia (2.28). Korzystając z (2.25) otrzymujemy:

$$\begin{aligned}
\sigma^2_{\Theta} S(\Theta) &= \int_{\mathcal{B}} dy P(y|\Theta) (S(\Theta) - E_{\Theta} S(\Theta))^2 \\
&= \int_{\mathcal{B}} dy P(y|\Theta) (S(\Theta))^2 = \int_{\mathcal{B}} dy P(y|\Theta) \left(\frac{\partial}{\partial \Theta} \ln P(y|\Theta) \right)^2 \\
&= \int_{\mathcal{B}} dy P(y|\Theta) \left[\left(\frac{\partial}{\partial \Theta} P(y|\Theta) \right) / P(y|\Theta) \right]^2 = \int_{\mathcal{B}} dy \left[\frac{\partial}{\partial \Theta} P(y|\Theta) \right]^2 / P(y|\Theta) .
\end{aligned} \tag{2.29}$$

Natomiast korzystając z (2.27) otrzymujemy:

$$\begin{aligned}
I_F(\Theta) &= E_{\Theta} \mathbf{IF}(\Theta) = \int_{\mathcal{B}} dy P(y|\Theta) \mathbf{IF}(\Theta) = - \int_{\mathcal{B}} dy P(y|\Theta) \left(\frac{\partial^2}{\partial \Theta^2} \ln P(y|\Theta) \right) \\
&= - \int_{\mathcal{B}} dy P(y|\Theta) \frac{\partial}{\partial \Theta} \left[\left(\frac{\partial}{\partial \Theta} P(y|\Theta) \right) / P(y|\Theta) \right] \\
&= \int_{\mathcal{B}} dy P(y|\Theta) \frac{\left[- \left(\frac{\partial^2}{\partial \Theta^2} P(y|\Theta) \right) P(y|\Theta) + \left(\frac{\partial}{\partial \Theta} P(y|\Theta) \right)^2 \right]}{(P(y|\Theta))^2} = \int_{\mathcal{B}} dy \left[\frac{\partial}{\partial \Theta} P(y|\Theta) \right]^2 / P(y|\Theta) ,
\end{aligned} \tag{2.30}$$

co porównując z (2.29) kończy dowód twierdzenia (2.28).

Wniosek: Dowodząc (2.25) okazało się, że przy założeniu warunku regularności, wartość oczekiwana na przestrzeni próby z funkcji wynikowej zeruje się, tzn. $E_{\Theta} S(\Theta) = 0$ i rezultat ten jest słuszny dla ogólnego przypadku wektorowego. Z (2.28) wynika również, że macierz wariancji-kowariancji funkcji wynikowej jest równa IF, tzn. $\sigma^2_{\Theta} S(\Theta) = I_F(\Theta)$, co jak sprawdzimy poniżej, ma znaczenie dla dodatniej określoności metryki Fishera g_{ij} w teorii pola. Skoro równości (2.25) oraz (2.28) zachodzą dla przypadku wektorowego, zatem są one również słuszne w przypadku skalarnym.

Ćwiczenie: Przedstawione dowody dla własności (2.25) oraz (2.28) były ogólne. Sprawdzić bezpośrednim rachunkiem, że zachodzą one dla (2.20) oraz (2.24) w powyższym przykładzie estymacji obu parametrów rozkładu normalnego.

Ważny związek: Z (2.29)-(2.30) widać, że przy spełnieniu wspomnianych własności regularności, zachodzi:

$$I_F(\Theta) = E_{\Theta} \mathbf{IF}(\Theta) = - \int_{\mathcal{B}} dy P(y|\Theta) \left(\frac{\partial^2}{\partial \Theta^2} \ln P(y|\Theta) \right) = \int_{\mathcal{B}} dy P(y|\Theta) \left[\frac{\partial}{\partial \Theta} \ln P(y|\Theta) \right]^2 . \tag{2.31}$$

Uwaga o dodatniej określoności obserwowanej IF: Niech g_{ij} są elementami $d \times d$ - wymiarowej macierzy $I_F(\Theta)$. Z (2.31) otrzymujemy dla dowolnego d - wymiarowego wektora $v = (v_1, v_2, \dots, v_d)^T$:

$$v^T I_F(\Theta) v = \sum_{i=1}^d \sum_{j=1}^d v_i g_{ij}(\Theta) v_j = \int_{\mathcal{B}} dy P(y|\Theta) \left(\sum_{i=1}^d v_i \frac{\partial \ln P(\Theta)}{\partial \theta_i} \right) \left(\sum_{j=1}^d \frac{\partial \ln P(\Theta)}{\partial \theta_j} v_j \right) \geq 0 , \tag{2.32}$$

co oznacza, że oczekiwana informacja Fishera $I_F(\Theta)$ jest *dodatnio półokreślona* [13], jak to zaznaczyliśmy poniżej (2.28). Jednak w teorii pola interesuje nas zaostrenie warunku (2.32) do własności dodatniej określoności.

Z (2.28) (lub (2.31)) widać również, że w teorii pola, którą dałoby się sformułować dla ciągłych, regularnych, unormowanych rozkładów, co pociąga za sobą ciągłość rozkładu funkcji wynikowej na całej przestrzeni próby \mathcal{B} , macierz $I_F(\Theta)$ jest określona dodatnio.

Ponieważ w Rozdziale 2.2 zwrócimy uwagę na fakt, że oczekiwana IF określa tzw. *metrykę Fishera-Rao* $(g_{ij}) := I_F(\Theta)$ na przestrzeni statystycznej \mathcal{S} , zatem:

W teorii pola z ciągłymi, regularnymi i unormowanymi rozkładami, metryka Fishera-Rao g_{ij} jest dodatnio określona.

2.2 Wstęp do geometrii różniczkowej na przestrzeni statystycznej i α -koneksja

Niech zbiorem punktów Ω reprezentujących konfigurację układu będzie przestrzeń próby układu \mathcal{B} . Np. w przypadku próby jednowymiarowej $\mathcal{B} \equiv \mathcal{Y}$, gdzie \mathcal{Y} jest zbiorem wszystkich możliwych wartości y zmiennej losowej Y . Niech P jest miarą probabilistyczną (prawdopodobieństwem) na \mathcal{B} . Zbiór wszystkich miar na \mathcal{B} oznaczmy $\Sigma(\mathcal{B})$ i nazwijmy *przestrzenią stanów* modelu.

Określenie modelu statystycznego: Rozważmy podzbiór $\mathcal{S} \subset \Sigma(\mathcal{B})$, na którym jest zadany układ współrzędnych $(\xi^i)_{i=1}^d$ [6] tak, że \mathcal{S} jest rozmaitością⁶. Niech na \mathcal{B} określony jest d - wymiarowy model statystyczny (tzn. para (y, P)), a precyzyjniej:

$$\mathcal{S} = \{P_{\Xi} \equiv P(y|\Xi), \Xi \equiv (\xi^i)_{i=1}^d \in V_{\Xi} \subset \mathbb{R}^d\}, \quad (2.34)$$

tzn. rodzina rozkładów prawdopodobieństwa parametryzowana przez d nielosowych zmiennych o wartościach rzeczywistych $(\xi^i)_{i=1}^d$ należących do przestrzeni parametru V_{Ξ} , będącej podzbiorem \mathbb{R}^d . Mówimy, że \mathcal{S} jest d - wymiarową przestrzenią statystyczną.

Notacja: Ponieważ w danym modelu statystycznym wartość parametru Ξ określa jednoznacznie rozkład prawdopodobieństwa P jako punkt na \mathcal{S} , więc ze względu na wygodę i o ile nie będzie to prowadziło do nieporozumień, sformułowania (punkt) $P_{\Xi} \in \mathcal{S}$ oraz (punkt) $\Xi \in \mathcal{S}$ będziemy stosowali zamiennie.

Uwaga o niezależności P od parametryzacji: Oczywiście rozkład prawdopodobieństwa nie zależy od wyboru bazy w przestrzeni statystycznej \mathcal{S} , tzn. gdyby np. Θ był innym układem współrzędnych (inną parametryzacją) to $P = P_{\Theta} \equiv P(y|\Theta) = P_{\Xi} \equiv P(y|\Xi)$ dla każdego $P \in \mathcal{S}$.

⁶ Niech \mathcal{S} jest zbiorem punktów, na którym określony jest układ współrzędnych $\phi_{\Xi} : \mathcal{S} \rightarrow \mathbb{R}^d$. **W skrypcie interesują nas tylko globalne układy współrzędnych.** Wtedy ϕ_{Ξ} odwzorowuje każdy punkt $P \in \mathcal{S}$ w zbiór d liczb rzeczywistych $\phi_{\Xi}(P) \equiv (\xi^1(P), \xi^2(P), \dots, \xi^d(P))^T = (\xi^1, \xi^2, \dots, \xi^d)^T \equiv \Xi$.

Niech istnieje zbiór układów współrzędnych (czyli atlas) \mathcal{A} spełniający następujące warunki:

- 1) Każdy element $\phi_{\Xi} = [(\xi^i)_{i=1}^d] \in \mathcal{A}$, jest wzajemnie jednoznacznym odwzorowaniem $\phi_{\Xi} : \mathcal{S} \rightarrow \mathbb{R}^d$ z \mathcal{S} w pewien otwarty podzbiór w \mathbb{R}^d .
- 2) Dla każdego $\phi_{\Xi} \in \mathcal{A}$ oraz wzajemnie jednoznacznego odwzorowania $\phi_{\Xi'} : \mathcal{S} \rightarrow \mathbb{R}^d$, zachodzi równoważność: $(\phi_{\Xi'} = (\xi'^i)_{i=1}^d \in \mathcal{A}) \Leftrightarrow (\phi_{\Xi'} \circ \phi_{\Xi}^{-1} \text{ jest dyfeomorfizmem rzędu } C^{\infty})$.

$$\text{Zbiór } \mathcal{S} \text{ z tak określonym atlasem } \mathcal{A} \text{ to } (C^{\infty} \text{ różniczkowalna}) \text{ rozmaitość}. \quad (2.33)$$

Określenie macierzy informacyjnej Fishera: Wprowadźmy oznaczenie:

$$\ell_{\Xi} \equiv l(y|\Xi) \equiv \ln P_{\Xi} \quad \text{oraz} \quad \partial_i \equiv \frac{\partial}{\partial \xi^i} \quad i = 1, 2, \dots, d. \quad (2.35)$$

Dla każdego punktu P_{Ξ} , $d \times d$ - wymiarowa macierz $(g_{ij}(\Xi))$ o elementach:

$$g_{ij}(\Xi) := E_{\Xi}(\partial_i \ell_{\Xi} \partial_j \ell_{\Xi}) = \int_{\mathcal{B}} dy P(y|\Xi) \partial_i l(y|\Xi) \partial_j l(y|\Xi), \quad i, j = 1, 2, \dots, d, \quad \forall P_{\Xi} \in \mathcal{S}, \quad (2.36)$$

jest nazywana *macierzą informacyjną Fishera* na \mathcal{S} w punkcie P_{Ξ} [6]. Wielkość $E_{\Xi}(\cdot)$ oznacza tutaj wartość oczekiwaną, a całkowanie przebiega po całej przestrzeni próby \mathcal{B} .

Przy założeniu spełnienia warunków regularności (porównaj (2.31)), macierz $(g_{ij}(\Xi))$ może zostać zapisana następująco:

$$g_{ij} = -E_{\Xi}(\partial_i \partial_j \ell_{\Xi}) = - \int_{\mathcal{B}} dy P(y|\Xi) \partial_i \partial_j \ln P(y|\Xi), \quad i, j = 1, 2, \dots, d, \quad \forall P_{\Xi} \in \mathcal{S}. \quad (2.37)$$

Metryka Fishera: Macierz informacyjna Fishera określa na \mathcal{S} iloczyn wewnętrzny nazywany *metryką Fishera* \langle, \rangle na \mathcal{S} , definiując go w układzie współrzędnych $(\xi^i)_{i=1}^d$ poprzez związek:

$$\langle \partial_i, \partial_j \rangle := g_{ij}, \quad \text{gdzie wektory bazowe} \quad \partial_i \equiv \frac{\partial}{\partial \xi^i}, \quad i = 1, 2, \dots, d, \quad \forall P_{\Xi} \in \mathcal{S}. \quad (2.38)$$

Metryka g_{ij} Fishera-Rao jest metryką typu Riemannowskiego⁷. Warto zauważyć, że na rozmaitości \mathcal{S} można zdefiniować nieskończoną liczbę metryk Riemannowskich. Jednak Chentsov pokazał, że *metryka Fishera-Rao jest wyróżniona spośród wszystkich innych tym, że (z dokładnością do stałego czynnika) jest ona jedyną, która jest redukowana* (w znaczeniu zmniejszania się odległości dowolnych dwóch stanów) *w każdym stochastycznym odwzorowaniu* [20, 21].

Określenie koneksji afinicznej: Oznaczmy przez T_P przestrzeń styczną do \mathcal{S} w punkcie $P_{\Xi} \in \mathcal{S}$. Każdy wektor $V \in T_P$ można rozłożyć na wektory bazowe $(\partial_i)_P$:

$$V = \sum_{i=1}^d V^i (\partial_i)_P, \quad \text{gdzie} \quad V \in T_P, \quad \forall P_{\Xi} \in \mathcal{S}, \quad (2.43)$$

⁷ **Określenie przestrzeni Riemannowskiej:** Niech \mathcal{S} jest rozmaitością i założmy, że dla każdego punktu $P_{\Xi} \in \mathcal{S}$ w jej przestrzeni stycznej $T_P \equiv T_P(\mathcal{S})$ jest określony iloczyn wewnętrzny \langle, \rangle_P taki, że $\langle V, W \rangle_P \in \mathbb{R}$ oraz posiadający dla dowolnych wektorów $V, W \in T_P$ następujące własności:

$$(\forall a, b \in \mathbb{R}) \quad \langle aV + bW, Z \rangle_P = a \langle V, Z \rangle_P + b \langle W, Z \rangle_P \quad \text{liniowość} \quad (2.39)$$

$$\langle V, W \rangle_P = \langle W, V \rangle_P \quad \text{symetryczność} \quad (2.40)$$

$$\text{Jeśli } V \neq 0, \text{ wtedy } \langle V, V \rangle_P > 0 \quad \text{dodatnia określoność.} \quad (2.41)$$

Pierwsze dwie własności oznaczają, że \langle, \rangle_P jest formą dwuliniową. Odwzorowanie $g : \mathcal{S} \ni P \rightarrow \langle, \rangle_P$ jest na \mathcal{S} polem wektorowym (kowariantnym) rangi 2. Nazywamy je *metryką Riemanna na \mathcal{S}* , a *przestrzeń \mathcal{S} , z tak określoną metryką g , nazywamy przestrzenią Riemannowską (\mathcal{S}, g)* .

Gdy dla każdego $P \in \mathcal{S}$ współrzędne $V^i, i = 1, 2, \dots, d$ dowolnego wektora $V \in T_P$ są C^∞ (tzn. są analityczne) względem pewnego układu współrzędnych Ξ , wtedy metryka Riemanna jest C^∞ . W skrypcie rozważamy tylko przypadek C^∞ .

Mając metrykę g możemy zdefiniować długość wektora stycznego V przez nią indukowaną:

$$\|V\| = \sqrt{\langle V, V \rangle_P} = \sqrt{\sum_{i,j=1}^d g_{ij}(\Xi) V^i V^j}. \quad (2.42)$$

gdzie dolny indeks P oznacza zależność układu współrzędnych $\partial_i \equiv \frac{\partial}{\partial \xi^i}$, $i = 1, 2, \dots, d$, od punktu $P_{\Xi} \in \mathcal{S}$.

Niech $\gamma_{PP'}$ oznacza ścieżkę w \mathcal{S} łączącą punkty P oraz P' . Przyporządkujemy, każdej ścieżce $\gamma_{PP'}$ w \mathcal{S} , odwzorowanie $\Pi_{\gamma_{PP'}}$:

$$\gamma_{PP'} \rightarrow \Pi_{\gamma_{PP'}} : T_P \rightarrow T_{P'}, \quad \forall P_{\Xi} \text{ i } P'_{\Xi} \in \mathcal{S}, \quad (2.44)$$

które przekształca wektory z przestrzeni wektorowej T_P do przestrzeni wektorowej $T_{P'}$.

Rozważmy trzy dowolne punkty P , P' oraz P'' w \mathcal{S} , oraz ścieżki $\gamma_{PP'}$ z punktu P do P' i $\gamma_{P'P''}$ z P' do P'' . Mówimy, że przekształcenie Π jest *koneksją afiniczną* jeśli:

$$\Pi_{\gamma_{PP''}} = \Pi_{\gamma_{P'P''}} \circ \Pi_{\gamma_{PP'}} \quad \text{oraz} \quad \Pi_{\gamma_0} \equiv \Pi_{\gamma_{PP}} = id, \quad \forall P_{\Xi} \text{ i } P'_{\Xi} \in \mathcal{S}, \quad (2.45)$$

gdzie id jest przekształceniem identycznościowym.

Aby określić postać liniowego odwzorowania $\Pi_{PP'}$ pomiędzy T_P a $T_{P'}$ musimy, dla każdego $i \in 1, 2, \dots, d$, określić $\Pi_{PP'}((\partial_i)_P)$ jako liniową kombinację wektorów bazowych $(\partial_1)_{P'}$, $(\partial_2)_{P'}$, ..., $(\partial_d)_{P'}$ w $T_{P'}$.

Określenie przesunięcia równoległego: Niech V_P będzie wektorem stycznym w P . Wektor $V_{PP'}$ nazywamy *równoległym przesunięciem wektora* V_P z P do P' wzdłuż krzywej (ścieżki) $\gamma_{PP'}$, wtedy gdy:

$$V_P \rightarrow V_{PP'} := \Pi_{\gamma_{PP'}} V_P, \quad \forall P_{\Xi} \text{ i } P'_{\Xi} \in \mathcal{S}. \quad (2.46)$$

Określenie równoległości wektorów: Dwa wektory V_P oraz $V_{P'}$ styczne do \mathcal{S} w punktach P i P' są *równoległe*, jeśli równoległe przesunięcie (określone w (2.46)) wzdłuż wskazanej krzywej γ jednego z nich, powiedzmy V_P , do punktu P' "zaczepienia" drugiego z nich, da wektor $V_{PP'}$, który jest *proporcjonalny* do $V_{P'}$ (i na odwrót).

Koneksja afiniczna pozwala zdefiniować pochodną kowariantną w następujący sposób. Niech V jest dowolnym wektorem w przestrzeni stycznej T_P , oraz niech $\gamma(t)$, gdzie $t \in \langle 0, 1 \rangle$, oznacza dowolną ścieżkę z P do P' w \mathcal{S} , która wychodzi z P w kierunku $W \in T_P$.

Pochodną kowariantną (koneksji afinicznej Π) wektora V w punkcie P i w kierunku $W \in T_P$ definiujemy następująco:

$$\nabla_W V := \frac{d}{dt} \left(\Pi_{\gamma_{P\gamma(t)}} V \right) \Big|_{t=0}, \quad \text{gdzie} \quad \gamma(t=0) = P, \quad \forall P_{\Xi} \in \mathcal{S}. \quad (2.47)$$

Pojęcie pochodnej kowariantnej pozwala wypowiedzieć się co do równoległości wektorów w punktach P i P' .

Określenie geodezyjnej: Geodezyjną nazywamy taką krzywą γ w \mathcal{S} , której wszystkie wektory styczne są do siebie równoległe. W związku z tym mówimy, że geodezyjna jest *samo-równoległą* krzywą w \mathcal{S} .

Współczynniki koneksji: Konkretna analityczna postać współczynników koneksji *musi być podana przy ustalonej parametryzacji*⁸ $\Xi \rightarrow P$, tzn. w określonym układzie współrzędnych $\Xi = (\xi^i)_{i=1}^d$. Zakładając, że różnica pomiędzy $\Pi_{PP'}((\partial_j)_P)$ oraz $(\partial_j)_{P'}$ jest *infinitesimalna*, i że może być wyrażona jako liniowa kombinacja *różniczek* $d\xi^1, d\xi^2, \dots, d\xi^d$, gdzie:

$$d\xi^i = \xi^i(P) - \xi^i(P'), \quad i = 1, 2, \dots, d, \quad P_\Xi \text{ i } P'_\Xi \in \mathcal{S}, \quad (2.50)$$

mamy:

$$\Pi_{PP'}((\partial_j)_P) = (\partial_j)_{P'} - \sum_{i,l=1}^d d\xi^i (\Gamma_{ij}^l)_P (\partial_l)_{P'}, \quad j = 1, 2, \dots, d, \quad P_\Xi \text{ i } P'_\Xi \in \mathcal{S}, \quad (2.51)$$

gdzie $(\Gamma_{ij}^l)_P$, $i, j, l = 1, 2, \dots, d$, są d^3 *współczynnikami koneksji* zależącymi od punktu P . Koneksję afiniczną Π , więc i współczynniki koneksji, można (przy ustalonej parametryzacji Ξ) określić na różne sposoby. Jeden z nich związany z koneksją Levi-Civita podany jest w (2.78). Jednak w analizie na przestrzeniach statystycznych szczególnie użyteczna okazała się tzw. α -koneksja.

α -koneksja: W każdym punkcie $P_\Xi \in \mathcal{S}$, α -koneksja zadaje d^3 funkcji $\Gamma_{ij,r}^{(\alpha)} : \Xi \rightarrow (\Gamma_{ij,r}^{(\alpha)})_\Xi$, $i, j, r = 1, 2, \dots, d$, przyporządkowując mu współczynniki koneksji o następującej postaci [6]:

$$(\Gamma_{ij,r}^{(\alpha)})_\Xi \equiv (\Gamma_{ij,r}^{(\alpha)})_{P_\Xi} = E_\Xi \left[\left(\partial_i \partial_j \ell_\Xi + \frac{1-\alpha}{2} \partial_i \ell_\Xi \partial_j \ell_\Xi \right) \partial_r \ell_\Xi \right], \quad \forall P_\Xi \in \mathcal{S}. \quad (2.52)$$

Koneksja α jest *symetryczna*, tzn.:

$$\Gamma_{ij,l}^{(\alpha)} = \Gamma_{ji,l}^{(\alpha)}, \quad i, j, l = 1, 2, \dots, d, \quad \forall P_\Xi \in \mathcal{S}. \quad (2.53)$$

W końcu, jeśli $g^{ij}(P)$ jest (i, j) -składową macierzy odwrotnej do macierzy informacyjnej $(g_{ij}(P))$, to współczynniki $\Gamma_{ij}^{r(\alpha)}$ są równe:

$$\Gamma_{ij}^{r(\alpha)} = \sum_{l=1}^d g^{rl} \Gamma_{ij,l}^{(\alpha)}, \quad i, j, r = 1, 2, \dots, d, \quad \forall P_\Xi \in \mathcal{S}. \quad (2.54)$$

⁸Niech $\Delta = (\delta^i)_{i=1}^d$ jest innym niż $\Xi = (\xi^i)_{i=1}^d$ układem współrzędnych (inną parametryzacją) i niech $\tilde{\partial}_j \equiv \frac{\partial}{\partial \delta^j} = \sum_{i=1}^d \frac{\partial \xi^i}{\partial \delta^j} \partial_i$. Można pokazać, że współczynniki koneksji w układach współrzędnych Δ oraz Ξ są związane ze sobą następująco [6]:

$$\tilde{\Gamma}_{rs}^t = \sum_{i,j,l=1}^d \left(\Gamma_{ij}^l \frac{\partial \xi^i}{\partial \delta^r} \frac{\partial \xi^j}{\partial \delta^s} + \frac{\partial^2 \xi^l}{\partial \delta^r \partial \delta^s} \right) \frac{\partial \delta^t}{\partial \xi^l}, \quad t, r, s = 1, 2, \dots, d. \quad (2.48)$$

Drugi składnik w (2.48) zależy tylko od postaci transformacji współrzędnych i jest niezależny od koneksji. Zatem, za wyjątkiem transformacji liniowej, dla której drugi składnik znika, współczynniki koneksji nie transformują się przy przejściu do nowego układu współrzędnych jak wielkość tensorowa. Gdyby więc w układzie współrzędnych $\Xi = (\xi^i)_{i=1}^d$ wszystkie $\Gamma_{ij}^l = 0$, to (za wyjątkiem transformacji liniowej) w nowym układzie współrzędnych $\Delta = (\delta^i)_{i=1}^d$ nie wszystkie współczynniki $\tilde{\Gamma}_{rs}^t$ byłyby również równe zero.

Jednakże postać koneksji afinicznej Π jest w nowej parametryzacji taka sama jak (2.51), tzn. jest ona współmiennicza, co oznacza, że:

$$\Pi_{PP'}((\tilde{\partial}_j)_P) = (\tilde{\partial}_j)_{P'} - \sum_{i,l=1}^d d\delta^i (\tilde{\Gamma}_{ij}^l)_P (\tilde{\partial}_l)_{P'}, \quad \text{gdzie } j = 1, 2, \dots, d \text{ oraz } d\delta^i = \delta^i(P) - \delta^i(P'). \quad (2.49)$$

α - pochodna kowariantna: Mając metrykę Fishera-Rao g_{ij} , (2.38), zdefiniujemy α - pochodną kowariantną $\nabla^{(\alpha)}$ na \mathcal{S} poprzez α -koneksję afiniczną następująco:

$$\langle \nabla_{\partial_i}^{(\alpha)} \partial_j, \partial_l \rangle := \Gamma_{ij,l}^{(\alpha)}, \quad i, j, l = 1, 2, \dots, d, \quad \forall P_{\Xi} \in \mathcal{S}. \quad (2.55)$$

Powyżej, poprzez koneksję afiniczną zdefiniowaliśmy pochodną kowariantną. Ale i odwrotnie, pochodna kowariantna definiuje koneksję [6]. Stąd np. mówimy, że określiliśmy koneksję ∇ .

Pole wektorowe na \mathcal{S} : Niech $V : P \rightarrow V_P$ jest odwzorowaniem przyporządkowującym każdemu punktowi $P \in \mathcal{S}$ wektor styczny $V_P \in T_P(\mathcal{S})$. Odwzorowanie to nazywamy *polem wektorowym*. Na przykład, jeśli $(\xi^i)_{i=1}^d$ jest układem współrzędnych, wtedy przyporządkowanie $\frac{\partial}{\partial \xi^i} : P \rightarrow \left(\frac{\partial}{\partial \xi^i} \right)_P$, $i = 1, 2, \dots, d$, określa d pól wektorowych na \mathcal{S} .

Oznaczmy przez $T(\mathcal{S})$ rodzinę wszystkich pól wektorowych klasy C^∞ typu $V_P = \sum_{i=1}^d V_P^i (\partial_i)_P \in T_P(\mathcal{S})$ na \mathcal{S} , gdzie d funkcji $V^i : P \rightarrow V_P^i$ nazywamy współrzędnymi pola wektorowego V ze względu na $(\xi^i)_{i=1}^d$.

Określenie pochodnej kowariantnej pola wektorowego: Rozważmy dwa pola wektorowe $V, W \in T(\mathcal{S})$. Niech w bazie $(\xi^i)_{i=1}^d$ pola te mają postać $V = \sum_{i=1}^d V^i \partial_i$ oraz $W = \sum_{i=1}^d W^i \partial_i$. Pochodną kowariantną pola V ze względu na W nazywamy pole wektorowe $\nabla_W V \in T(\mathcal{S})$, które w bazie $(\xi^i)_{i=1}^d$ ma postać:

$$\nabla_W V = \sum_{i=1}^d W^i \sum_{k=1}^d \left\{ \partial_i V^k + \sum_{j=1}^d V^j \Gamma_{ij}^k \right\} \partial_k, \quad \forall P_{\Xi} \in \mathcal{S}. \quad (2.56)$$

Przyjmując $W = \partial_i$ oraz $V = \partial_j$ łatwo sprawdzić, że związek (2.56) daje:

$$\nabla_{\partial_i} \partial_j = \sum_{k=1}^d \Gamma_{ij}^k \partial_k, \quad \forall P_{\Xi} \in \mathcal{S}, \quad (2.57)$$

co po skorzystaniu z (2.54) jest zgodne z określeniem koneksji ∇ w (2.55).

Można pokazać [6], że zachodzi:

$$\nabla^{(\alpha)} = \frac{1+\alpha}{2} \nabla^{(1)} + \frac{1-\alpha}{2} \nabla^{(-1)}, \quad \forall P_{\Xi} \in \mathcal{S}. \quad (2.58)$$

Uwaga o płaskości \mathcal{S} ze względu na ∇ : Mówimy, że układ współrzędnych $(\xi^i)_{i=1}^d$ jest *afinicznym układem współrzędnych dla koneksji ∇* gdy zachodzą (równoważne) warunki:

$$\nabla_{\partial_i} \partial_j = 0 \quad \text{lub równoważnie} \quad \Gamma_{ij}^l = 0 \quad i, j, l = 1, 2, \dots, d, \quad \forall P_{\Xi} \in \mathcal{S}. \quad (2.59)$$

Jeśli dla zadanej koneksji ∇ odpowiadający jej układ współrzędnych jest afiniczny, to mówimy, że koneksja ∇ jest płaska⁹ lub, że *przestrzeń statystyczna \mathcal{S} jest płaska ze względu na ∇* .

⁹W ogólności, znikanie koneksji jest warunkiem wystarczającym lecz niekoniecznym afinicznej płaskości, która w bardziej fundamentalnym określeniu ma miejsce, gdy tensor krzywizny (Riemanna) R zeruje się na całej rozmaitości \mathcal{S} , tzn. zerują się wszystkie jego składowe. Składowe tensora krzywizny R w bazie $(\xi^i)_{i=1}^d$ mają następującą postać:

$$R_{ijk}^l \equiv R_{ijk}^l(\Gamma) = \partial_i \Gamma_{jk}^l - \partial_j \Gamma_{ik}^l + \sum_{s=1}^d \Gamma_{is}^l \Gamma_{jk}^s - \sum_{s=1}^d \Gamma_{js}^l \Gamma_{ik}^s, \quad \forall P_{\Xi} \in \mathcal{S}. \quad (2.60)$$

Gdy na rozmaitości \mathcal{S} istnieje globalny układ współrzędnych, wtedy dla koneksji, która jest symetryczna (tak jak np. α -koneksja), znikanie tensora krzywizny pociąga istnienie układu współrzędnych, w którym znika koneksja i oba określenia afinicznej płaskości pokrywają się.

Płaskość \mathcal{S} ze względu na ∇ oznacza, że wszystkie wektory bazowe $\partial_i \equiv \frac{\partial}{\partial \xi^i}$, $i = 1, 2, \dots, d$ są równoległe na całej przestrzeni \mathcal{S} .

Modele z koneksją $\nabla^{(\alpha)}$, które są $\alpha = +1$ bądź -1 płaskie, odgrywają szczególną rolę w modelowaniu statystycznym [13].

Dwa przykłady rodzin rozkładów prawdopodobieństwa: Istnieją dwa przypadki rodzin modeli statystycznych, szczególnie istotnych w badaniu podstawowych geometrycznych własności modeli statystycznych. Pierwsza z nich to rodzina rozkładów eksponentialnych, a druga, rozkładów mieszanych.

Rodzina modeli eksponentialnych okazuje się wyjątkowo ważna, nie tylko dla badania własności statystycznych, ale również w związku z jej realizacją w szeregu zagadnieniach fizycznych. Niech wielkość próby $N = 1$. Niech zmienna losowa Y będzie zmienną typu ciągłego lub dyskretnego. Ogólna postać regularnej rodziny rozkładów eksponentialnych¹⁰ jest następująca:

$$p_{\Xi} \equiv p(\mathbf{y}|\Xi) = \exp \left[C(\mathbf{y}) + \sum_{i=1}^d \xi^i F_i(\mathbf{y}) - \psi(\Xi) \right] = \exp \left[\sum_{i=1}^d \xi^i F_i(\mathbf{y}) - \psi(\Xi) \right] h(\mathbf{y}) , \quad (2.61)$$

gdzie ξ^i , $i = 1, 2, \dots, d$, są tzw. *parametrami kanonicznymi*, natomiast $h(\mathbf{y}) = \exp(C(\mathbf{y}))$ jest nieujemną funkcją, która nie zależy od wektorowego parametru Ξ .

W (2.61) pojawiła się d -wymiarowa statystyka:

$$F(Y) = (F_1(Y), \dots, F_d(Y))^T \equiv (F_i(Y))_{i=1}^d, \quad (2.62)$$

nazywana statystyką *kanoniczną*.

Całkując obustronnie (2.61) po przestrzeni próby $\mathcal{B} = \mathcal{Y}$, a następnie wykorzystując własność normalizacji $\int_{\mathcal{Y}} d\mathbf{y} p(\mathbf{y}|\Xi) = 1$, otrzymujemy¹¹:

$$\psi(\Xi) = \ln \int_{\mathcal{Y}} d\mathbf{y} \exp \left[C(\mathbf{y}) + \sum_{i=1}^d \xi^i F_i(\mathbf{y}) \right] . \quad (2.63)$$

Modele eksponentialne są $\alpha = 1$ - płaskie, co oznacza, że [6]:

$$(\Gamma_{ij,k}^{(1)})_{\Xi} = 0 , \quad \text{dla modeli eksponentialnych} , \quad \forall P_{\Xi} \in \mathcal{S} . \quad (2.64)$$

Istotnie, korzystając z oznaczenia wprowadzonego w (2.35), otrzymujemy dla modeli zadanych przez (2.61):

$$\frac{\partial l(\mathbf{y}|\Xi)}{\partial \xi^i} = F_i(\mathbf{y}) - \frac{\partial \psi(\Xi)}{\partial \xi^i} , \quad (2.65)$$

¹⁰Zlogarytmowane modele eksponentialne $\{\ln p_{\Xi}, \Xi \equiv (\xi^i)_{i=1}^d \in V_{\Xi} \subset \mathbb{R}^d\}$ są przestrzeniami afinicznymi. Dlatego z punktu widzenia ich geometrycznej charakterystyki odgrywają taką rolę jak proste i powierzchnie w 3-wymiarowej geometrii Euklidesowej.

¹¹Funkcja $\psi(\Xi)$ okazuje się być tzw. potencjałem transformacji Legendre'a pomiędzy afinicznymi układami współrzędnych modeli eksponentialnych (por. Rozdziały 2.2.2.1 oraz 2.8.1).

skąd (przy okazji) otrzymujemy obserwowaną informację Fishera dla rozkładów eksponentialnych w parametryzacji kanonicznej:

$$\mathbf{iF}(\Xi) = -\frac{\partial^2 l(\mathbf{y}|\Xi)}{\partial \xi^j \partial \xi^i} = \frac{\partial^2 \psi(\Xi)}{\partial \xi^j \partial \xi^i} . \quad (2.66)$$

Podstawiając (2.66) do (2.52), otrzymujemy dla $\alpha = 1$ współczynniki 1-koneksji, równe:

$$(\Gamma_{ij,r}^{(1)})_{\Xi} = E_{\Xi} \left[\left(\frac{\partial^2 l(Y|\Xi)}{\partial \xi^j \partial \xi^i} \right) \frac{\partial l(Y|\Xi)}{\partial \xi_r} \right] = -\frac{\partial^2 \psi(\Xi)}{\partial \xi^j \partial \xi^i} E_{\Xi} \left[\frac{\partial l(Y|\Xi)}{\partial \xi_r} \right] = 0 , \quad \forall P_{\Xi} \in \mathcal{S} , \quad (2.67)$$

gdzie w ostatniej równości skorzystano z (2.25) dla funkcji wynikowej.

Ponadto dla modeli eksponentialnych z (2.37) oraz (2.66) otrzymujemy następującą postać metryki Fishera-Rao:

$$g_{ij} = -E_{\Xi}(\partial_i \partial_j \ell_{\Xi}) = \frac{\partial^2 \psi(\Xi)}{\partial \xi^j \partial \xi^i} . \quad (2.68)$$

Zwróćmy uwagę na równość prawych stron (2.66) oraz (2.68) dla rozkładów eksponentialnych.

Przykładami modeli z eksponentialnej rodziny rozkładów są:

i) *Rozkład normalny*, (1.16), $p(Y = \mathbf{y}|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{(\mathbf{y}-\mu)^2}{2\sigma^2}\right)$, $\mathbf{y} \in \mathbf{R}$, dla którego:

$$\begin{aligned} C(\mathbf{y}) &= 0 , \quad F_1(\mathbf{y}) = \mathbf{y} , \quad F_2(\mathbf{y}) = \mathbf{y}^2 , \quad \xi^1 = \frac{\mu}{\sigma^2} , \quad \xi^2 = -\frac{1}{2\sigma^2} , \\ \psi(\xi) &= \frac{\mu^2}{2\sigma^2} + \ln(\sqrt{2\pi}\sigma) = -\frac{(\xi^1)^2}{4\xi^2} + \frac{1}{2} \ln\left(-\frac{\pi}{\xi^2}\right) . \end{aligned} \quad (2.69)$$

ii) *Rozkład Poissona* (1.33), $p(Y = \mathbf{y}|\mu) = \frac{\mu^{\mathbf{y}} \exp(-\mu)}{\mathbf{y}!}$, gdzie $\mathbf{y} = 0, 1, \dots, \infty$, dla którego:

$$C(\mathbf{y}) = -\ln(\mathbf{y}!) , \quad F(\mathbf{y}) = \mathbf{y} , \quad \xi = \ln \mu , \quad \psi(\xi) = \mu = \exp \xi . \quad (2.70)$$

iii) *(Standardowy) rozkład eksponentialny*, $p(Y = \mathbf{y}|\mu) = \mu^{-1} \exp(-\mathbf{y}/\mu)$, gdzie $\mathbf{y} > 0$, dla którego:

$$C(\mathbf{y}) = 0 , \quad F(\mathbf{y}) = \mathbf{y} , \quad \xi = -\frac{1}{\mu} , \quad \psi(\xi) = \ln \mu = \ln\left(-\frac{1}{\xi}\right) . \quad (2.71)$$

Modele eksponentialne wyróżnia fakt osiągnięcia dolnego ograniczenia nierówności Rao-Cramera [21] (porównaj związek (2.18)).

Wymiar statystyki dostatecznej dla (2.61): Dla modeli eksponentialnych zachodzi ważna własność związana z wymiarem statystyki kanonicznej $F(Y)$, (2.62). Rozważmy N -elementową próbę $\tilde{Y} \equiv (Y_n)_{n=1}^N$, dla której każdy punktowy rozkład ma postać (2.61). Wtedy funkcja wiarygodności dla próby jest następująca:

$$P(\mathbf{y}|\Xi) = \exp \left[\sum_{i=1}^d \xi^i \sum_{n=1}^N F_i(\mathbf{y}_n) - N \psi(\Xi) \right] \exp \left[\sum_{n=1}^N C(\mathbf{y}_n) \right] . \quad (2.72)$$

Dostateczna statystyka dla wektorowego parametru oczekiwanego $(\theta_i)_{i=1}^d$, gdzie $\theta_i = E_{\Xi} [F_i(Y)]$, $i = 1, 2, \dots, d$ (por. (2.237) w Rozdziale 2.8.1) ma zatem postać:

$$\left(\sum_{n=1}^N F_1(\mathbf{y}_n), \sum_{n=1}^N F_2(\mathbf{y}_n), \dots, \sum_{n=1}^N F_d(\mathbf{y}_n) \right). \quad (2.73)$$

Jej wymiar jest równy d i jak widać, dla konkretnych reprezentantów ogólnej rodziny eksponentialnej nie zależy on od wymiaru próby N . Własność ta nie jest spełniona np. dla takich nieeksponentialnych rozkładów jak rozkładu Weibull'a oraz Pareto [11], dla których wymiar statystyki dostatecznej rośnie wraz z wymiarem próby. Inną ważną własnością rozkładów nieeksponentialnych jest to, że dziedzina (tzn. nośnik) ich funkcji gęstości może zależeć od parametru.

Rodzina mieszanych rozkładów prawdopodobieństwa:

$$p_{\Delta} \equiv p(\mathbf{y}|\Delta) = C(\mathbf{y}) + \sum_{i=1}^d \delta^i F_i(\mathbf{y}), \quad (2.74)$$

gdzie δ^i są tzw. *parametrami mieszanymi*.

Ćwiczenie: Pokazać, że przestrzeń statystyczna rodziny rozkładów mieszanych jest $\alpha = -1$ - płaska [6]:

$$(\Gamma_{ij,k}^{(-1)})_{\Xi} = 0, \quad \text{dla modeli mieszanych, } \forall P_{\Xi} \in \mathcal{S}. \quad (2.75)$$

Geometria przestrzeni \mathcal{S} a parametryzacja: Jako własności geometryczne przestrzeni \mathcal{S} przyjmujemy te, które są niezmiennicze ze względu na zmianę parametryzacji.

Np. własności geometryczne rodziny modeli eksponentialnych nie zależą od tego czy posłużymy się parametrami kanonicznymi $(\xi^i)_{i=1}^d$ czy oczekiwanymi $\theta_i = E_{\Xi} [F_i(Y)]$, $i = 1, 2, \dots, d$ (por. (2.237)).

2.2.1 Przestrzeń statystyczna dualnie płaska

Z (2.36) oraz (2.52) wynika, że:

$$\partial_r g_{ij} = \Gamma_{ri,j}^{(0)} + \Gamma_{rj,i}^{(0)}, \quad \forall P_{\Xi} \in \mathcal{S}, \quad (2.76)$$

co oznacza, że 0-koneksja jest metryczna¹² ze względu na metrykę Fishera-Rao [6].

¹²**Koneksja metryczna:** Poprzez metrykę Riemannowską g , której przykładem jest metryka Fishera-Rao, (por. Rozdział 2.2), można określić koneksję $\gamma_{PP'} \rightarrow \Pi_{\gamma_{PP'}}$, która jest metryczna na \mathcal{S} . Z *koneksją metryczną* mamy do czynienia gdy:

$$g_{P'}(\Pi_{\gamma_{PP'}} V, \Pi_{\gamma_{PP'}} W) = g_P(V, W), \quad \forall P_{\Xi} \text{ i } P'_{\Xi} \in \mathcal{S}, \quad (2.77)$$

tzn. gdy *iloczyn wewnętrzny jest niezmienniczy ze względu na przesunięcie równoległe* dla wszystkich wektorów stycznych V oraz W i wszystkich ścieżek γ z P do P' .

Koneksję metryczną nazywamy *koneksją Levi-Civita* (lub Riemannowską), jeśli jest ona zarówno metryczna jak i symetryczna. Dla zadanej g koneksja taka jest określona jednoznacznie jako:

$$\Gamma_{ij,k} = \frac{1}{2} (\partial_i g_{jk} + \partial_j g_{ki} - \partial_k g_{ij}), \quad \forall P_{\Xi} \in \mathcal{S}, \quad (2.78)$$

gdzie jej *symetria* oznacza spełnienie warunku (2.53). Koneksja (2.78) spełnia warunek (2.76).

Jednak w ogólności dla $\alpha \neq 0$, α -koneksja afiniczna nie jest metryczna, natomiast spełnia warunek dualności omówiony poniżej.

Koneksje dualne: Niech na rozmaitości \mathcal{S} zadana jest pewna metryka Riemannowska $g = \langle, \rangle$ i dwie koneksje ∇ oraz ∇^* . Metryka ta może być np. metryką g Fishera-Rao. Jeśli dla wszystkich pól wektorowych $V, W, Z \in T(\mathcal{S})$, zachodzi:

$$Z \langle V, W \rangle = \langle \nabla_Z V, W \rangle + \langle V, \nabla_Z^* W \rangle, \quad \forall P_\Xi \in \mathcal{S}, \quad (2.80)$$

wtedy mówimy, że ∇ oraz ∇^* są ze względu na metrykę \langle, \rangle *dualne* (sprzężone) względem siebie. W układzie współrzędnych $(\xi^i)_{i=1}^d$ metryka g ma współrzędne g_{ij} , a koneksje ∇ oraz ∇^* mają współczynniki koneksji odpowiednio $\Gamma_{ij,r}$ oraz $\Gamma_{ij,r}^*$.

Warunek dualności (2.80) możemy teraz zapisać w postaci¹³:

$$\partial_r g_{ij} = \Gamma_{ri,j} + \Gamma_{rj,i}^*, \quad \forall P_\Xi \in \mathcal{S}, \quad (2.82)$$

będące uogólnieniem warunku (2.76) istniejącego dla koneksji metrycznej.

Ponadto, koneksja $\nabla^{met} \equiv (\nabla + \nabla^*)/2$ jest koneksją metryczną, dla której zachodzi warunek $\partial_r g_{ij} = \Gamma_{ri,j}^{met} + \Gamma_{rj,i}^{met}$ [6]. Można również sprawdzić, że (ze względu na g) zachodzi $(\nabla^*)^* = \nabla$.

Struktura dualna: Trójkę $(g, \nabla, \nabla^*) \equiv (\mathcal{S}, g, \nabla, \nabla^*)$ nazywamy *strukturą dualną na \mathcal{S}* . W ogólności, mając metrykę g oraz koneksję ∇ określoną na \mathcal{S} , koneksja dualna ∇^* jest wyznaczona w sposób jednoznaczny, co jest treścią poniższego twierdzenia.

Twierdzenie o związku pomiędzy koneksjami dualnymi: Niech P oraz P' są punktami brzegowymi ścieżki γ , oraz niech przekształcenia $\Pi_{\gamma_{PP'}}$ oraz $\Pi_{\gamma_{PP'}}^*$ z $T_P(\mathcal{S})$ do $T_{P'}(\mathcal{S})$ opisują *równoległe przesunięcie wzdłuż γ* , odpowiednio ze względu na koneksje afiniczne ∇ oraz ∇^* . Wtedy dla wszystkich $V, W \in T_P(\mathcal{S})$ zachodzi [6]:

$$g_{P'}(\Pi_{\gamma_{PP'}} V, \Pi_{\gamma_{PP'}}^* W) = g_P(V, W), \quad \forall P_\Xi \text{ i } P'_\Xi \in \mathcal{S}. \quad (2.83)$$

Warunek ten jest uogólnieniem warunku (2.77) istniejącego dla koneksji metrycznej o *niezmienniczości iloczynu wewnętrznego ze względu na przesunięcie równoległe*. Wyznacza on w sposób jednoznaczny związek pomiędzy $\Pi_{\gamma_{PP'}}$ oraz $\Pi_{\gamma_{PP'}}^*$.

Niezmienniczość iloczynu wewnętrznego dla płaskich koneksji dualnych przy przesunięciu równoległym: Jeśli $\Pi_{\gamma_{PP'}}$ nie zależy na \mathcal{S} od ścieżki γ , a tylko od punktów końcowych¹⁴ P oraz P' , wtedy

Geodezyjne koneksji Levi-Civita są krzywymi o najmniejszej (mierzonej przez metrykę g) długości, tzn. jej długość pomiędzy punktem P a P' określona następująco:

$$\|\gamma\| = \int_P^{P'} \left\| \frac{d\gamma}{dt} \right\| dt = \int_P^{P'} \sqrt{\sum_{i,j=1}^d g_{ij} \frac{d\gamma^i}{dt} \frac{d\gamma^j}{dt}} dt, \quad (2.79)$$

jest najmniejsza, gdzie $\gamma^i(t)$ jest i -tą współrzędną punktu na krzywej $\gamma(t)$, tzn. $\gamma^i(t) := \xi^i(\gamma(t))$.

¹³ Wystarczy zapisać (2.80) dla wektorów bazowych $(\partial_i)_{i=1}^d$, otrzymując:

$$\partial_r \langle \partial_i, \partial_j \rangle = \langle \nabla_{\partial_r} \partial_i, \partial_j \rangle + \langle \partial_i, \nabla_{\partial_r}^* \partial_j \rangle, \quad \forall P_\Xi \in \mathcal{S}, \quad (2.81)$$

skąd wykorzystując (2.55) dla koneksji ∇ oraz ∇^* , otrzymujemy (2.82).

¹⁴ Mówimy wtedy, że koneksja na \mathcal{S} jest całkowalna. Gdy przestrzeń \mathcal{S} jest jedno-spójna, to warunek ten oznacza znikanie tensora krzywizny Riemanna na \mathcal{S} .

$\Pi_{\gamma_{PP'}} = \Pi_{PP'}$ na \mathcal{S} . Ponieważ przy określonej metryce g , koneksja dualna ∇^* jest jednoznacznie wyznaczona dla ∇ , zatem również dla koneksji ∇^* jest na \mathcal{S} spełniony warunek $\Pi_{\gamma_{PP'}}^* = \Pi_{PP'}^*$, skąd z (2.83) przy przesunięciu równoległym otrzymujemy:

$$g_{P'}(\Pi_{PP'}V, \Pi_{PP'}^*W) = g_P(V, W), \quad \forall P_{\Xi} \text{ i } P'_{\Xi} \in \mathcal{S}. \quad (2.84)$$

Zadanie: Korzystając z metryki (2.68) modelu eksponentialnego oraz z (2.67) i (2.52), sprawdzić bezpośrednim rachunkiem w parametryzacji kanonicznej $(\xi^i)_{i=1}^d$ warunek (2.82), otrzymując:

$$\partial_r g_{ij} = \Gamma_{ri,j}^{+1} + \Gamma_{rj,i}^{-1} = \Gamma_{rj,i}^{-1} = \partial_r \partial_i \partial_j \psi(\Xi), \quad \forall P_{\Xi} \in \mathcal{S}. \quad (2.85)$$

2.2.2 Dualne układy współrzędnych

Określenie dualnie płaskiej przestrzeni: Mówimy, że $(\mathcal{S}, g, \nabla, \nabla^*)$ jest *dualnie płaską przestrzenią*, jeśli obie koneksje dualne, ∇ oraz ∇^* , są płaskie na \mathcal{S} . Oznacza to, że jeśli koneksja ∇ jest płaska w pewnej bazie $(\xi^i)_{i=1}^d$ to koneksja ∇^* jest płaska w pewnej bazie $(\xi^{*i})_{i=1}^d$, którą nazywamy bazą dualną do $(\xi^i)_{i=1}^d$.

α - koneksja dualnie płaska:

Istotność pojęcia α -koneksji pojawia się wraz z rozważeniem na przestrzeni statystycznej \mathcal{S} nie tyle prostej pary $(g, \nabla^{(\alpha)}) \equiv (\mathcal{S}, g, \nabla^{(\alpha)})$ ale struktury potrójnej $(g, \nabla^{(\alpha)}, \nabla^{(-\alpha)}) \equiv (\mathcal{S}, g, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$. Powodem jest istnienie poprzez metrykę Fishera-Rao *dualności pomiędzy koneksjami* $\nabla^{(\alpha)}$ oraz $\nabla^{(-\alpha)}$, która okazuje się być ważna przy badaniu modeli statystycznych.

Podsumowując, dla dowolnego modelu statystycznego \mathcal{S} , zachodzi [6]:

$$(\mathcal{S} \text{ jest } \alpha - \text{płaska}) \Leftrightarrow (\mathcal{S} \text{ jest } (-\alpha) - \text{płaska}). \quad (2.86)$$

Przykład: Model statystyczny eksponentialny \mathcal{S} jest płaski w parametryzacji kanonicznej $(\xi^i)_{i=1}^d$ dla $\alpha = 1$. Istnieje zatem *parametryzacja dualna*, w której jest on również $\alpha = -1$ płaski. Np. w Rozdziale 2.8.1 okaże się, że dla modeli eksponentialnych spełniających warunek maksymalizacji entropii, bazą dualną do bazy kanonicznej jest baza parametrów oczekiwanych.

Podobnie jest dla rodziny rozkładów mieszanych, tzn. jest ona jednocześnie ± 1 płaska.

Określenie dualnych układów współrzędnych: Zastanówmy się nad ogólną strukturą przestrzeni $(\mathcal{S}, g, \nabla, \nabla^*)$, która byłaby dualnie płaska.

Z określenia przestrzeni płaskiej ze względu na określoną koneksję (2.59) oraz z (2.86) wynika, że jeśli istnieje układ współrzędnych $\Xi = (\xi^i)_{i=1}^d$ z wektorami bazowymi $\partial_{\xi^i} \equiv \frac{\partial}{\partial \xi^i}$, ze względu na który koneksja ∇ jest płaska, tzn. $\nabla_{\partial_{\xi^i}} \partial_{\xi^j} = 0$, $i, j = 1, 2, \dots, d$, to istnieje również układ współrzędnych $\Theta = (\theta^i)_{i=1}^d := (\xi^{*i})_{i=1}^d$ z wektorami bazowymi $\partial_{\theta^i} \equiv \frac{\partial}{\partial \theta^i}$, ze względu na który koneksja ∇^* jest płaska, tzn. $\nabla_{\partial_{\theta^j}} \partial_{\theta^i} = 0$, $j, l = 1, 2, \dots, d$.

Wniosek: Zatem, gdy pole wektorowe ∂_{ξ^i} jest ∇ -płaskie, więc pole wektorowe ∂_{θ^j} jest ∇^* -płaskie i z (2.84) wynika stałość $g_{P_{\Xi}}(\partial_{\xi^i}, \partial_{\theta^j})$ na \mathcal{S} . Fakt ten, biorąc pod uwagę wszystkie d stopni swobody

zawarte w afinicznym układzie współrzędnych, (2.59), można zapisać jako¹⁵:

$$\langle \partial_{\xi^i}, \partial_{\theta^j} \rangle \equiv g_{P_{\Xi}} \left(\frac{\partial}{\partial \xi^i}, \frac{\partial}{\partial \theta^j} \right) = \delta_{ij}, \quad i, j = 1, 2, \dots, d, \quad \forall P_{\Xi} \in \mathcal{S}, \quad (2.88)$$

gdzie iloczyn wewnętrzny $\langle \cdot, \cdot \rangle$ jest wyznaczony w konkretnym układzie współrzędnych, w tym przypadku w $\Xi = (\xi^i)_{i=1}^d$. Układy współrzędnych $\Xi = (\xi^i)_{i=1}^d$ oraz $\Theta = (\theta^j)_{j=1}^d$ określone na przestrzeni Riemannowskiej (\mathcal{S}, g) i spełniające warunek (2.88) nazywamy *wzajemnie dualnymi*. Warunek (2.88) oznacza stałość na \mathcal{S} iloczynu wewnętrznego dla układów dualnie płaskich.

W ogólności dla dowolnej przestrzeni Riemannowskiej (\mathcal{S}, g) nie istnieją układy współrzędnych wzajemnie dualne.

Jeśli jednak przestrzeń Riemannowska z dualną koneksją $(\mathcal{S}, g, \nabla, \nabla^*)$ jest dualnie płaska, to taka para układów współrzędnych *istnieje*. Ale i na odwrót. Jeśli na przestrzeni Riemannowskiej (\mathcal{S}, g) istnieją dwa układy współrzędnych $(\xi^i)_{i=1}^d$ oraz $(\theta^j)_{j=1}^d$ spełniające warunek (2.88), wtedy koneksje ∇ oraz ∇^* , względem których układy te są afiniczne, są określone, a $(\mathcal{S}, g, \nabla, \nabla^*)$ jest dualnie płaska.

¹⁵ Niech $\Xi = (\xi^i)_{i=1}^d$ będzie układem współrzędnych afinicznych. Wtedy, zgodnie z (2.48) współczynniki koneksji w układzie współrzędnych $\Theta = (\theta^i)_{i=1}^d$ są równe:

$$\tilde{\Gamma}_{rs}^t = \sum_{i,j,l=1}^d \left(\frac{\partial^2 \xi^l}{\partial \theta^r \partial \theta^s} \right) \frac{\partial \theta^t}{\partial \xi^i}, \quad t, r, s = 1, 2, \dots, d.$$

Zażądajmy aby nowy układ współrzędnych Θ był również afiniczny. Koniecznym i wystarczającym warunkiem spełnienia tego żądania jest zerowanie się wszystkich drugich pochodnych $\frac{\partial^2 \xi^l}{\partial \theta^r \partial \theta^s} = 0$, gdzie $l, r, s = 1, 2, \dots, d$. Warunek ten oznacza, że pomiędzy współrzędnymi $(\xi^i)_{i=1}^d$ oraz $(\theta^i)_{i=1}^d$ istnieje *afiniczna transformacja*:

$$\xi^l \longrightarrow \theta^l = \sum_{k=1}^d A^l_k \xi^k + V^l, \quad l = 1, 2, \dots, d, \quad \forall P \in \mathcal{S},$$

gdzie $A = (A^l_k) = (\partial \theta^l / \partial \xi^k) \in GL(d)$ jest niezależną od współrzędnych $d \times d$ -wymiarową nieosobliwą macierzą ogólnej grupy liniowych transformacji, natomiast jest $V = \sum_{l=1}^d V^l \partial_{\xi^l}$ jest stałym d -wymiarowym wektorem.

Metryka g jest tensorem kowariantnym rangi 2, jest to więc forma dwu-liniowa, zatem iloczyn wewnętrzny wektorów układów dualnych $\partial / \partial \xi^s$ oraz $\partial / \partial \theta^r$, można zapisać następująco:

$$\begin{aligned} \langle \partial / \partial \xi^s, \partial / \partial \theta^r \rangle &= \left\langle \partial / \partial \xi^s, \sum_{l=1}^d \frac{\partial \xi^l}{\partial \theta^r} \frac{\partial}{\partial \xi^l} \right\rangle = \sum_{l=1}^d \frac{\partial \xi^l}{\partial \theta^r} \langle \partial / \partial \xi^s, \partial / \partial \xi^l \rangle \\ &= \sum_{l=1}^d (A^{-1})^l_r g_{sl} = \delta_{sr} \quad s, r = 1, 2, \dots, d, \quad \forall P \in \mathcal{S}, \end{aligned} \quad (2.87)$$

gdzie $g_{sl} = \langle \partial / \partial \xi^s, \partial / \partial \xi^l \rangle$, a ostatnia równość w (2.87) wynika z żądania stałości iloczynu wewnętrznego $\langle \partial / \partial \xi^s, \partial / \partial \theta^r \rangle$ na \mathcal{S} dla układów dualnie płaskich zgodnie z (2.59) oraz z nieosobliwości transformacji afinicznej, tak że układ współrzędnych Θ ma tyle samo stopni swobody co układ Ξ . Zatem otrzymaliśmy warunek (2.88) dla układów dualnie płaskich.

Przechodząc zgodnie z (2.92) od współrzędnych kontrawariantnych θ^r do kowariantnych, θ_r , $r = 1, 2, \dots, d$, widać, że powyższy związek (2.87) dla wektorów bazowych układów dualnych można zapisać w postaci $\langle \partial / \partial \xi^s, \partial / \partial \theta_r \rangle = \delta_s^r$ podanej w [6].

Uwaga: Przestrzeń wektorów dualnych $(\partial / \partial \theta^r)_{r=1}^d$ układu, który jest afiniczny względem koneksji ∇^* jest izomorficzna z przestrzenią 1-form $(d\theta^r)_{r=1}^d$ sprzężoną liniowo do układu przestrzeni wektorów $(\partial / \partial \theta^r)_{r=1}^d$. Dla 1-form sprzężonych do wektorów bazowych z definicji zachodzi $d\theta^r(\partial / \partial \theta^s) = \delta_s^r$, $s, r = 1, 2, \dots, d$. [1-form $d\theta^r$, $r = 1, 2, \dots, d$, nie należy mylić z przyrostami typu (2.50) czy jak w (2.49)].

Euklidesowy układ współrzędnych: W przypadku Euklidesowego układu współrzędnych na \mathcal{S} mamy (z definicji):

$$\langle \partial_{\xi^i}, \partial_{\xi^j} \rangle = \delta_{ij} , \quad i, j = 1, 2, \dots, d \quad , \quad \forall P_{\Xi} \in \mathcal{S} , \quad (2.89)$$

co oznacza, że jest on *samo-dualny*.

Uwaga o współczesnym zastosowaniu koneksji dualnych: Interesującym wydaje się fakt, że pojęcie koneksji dualnych ma coraz większe zastosowanie w analizie układów liniowych [22] i szeregów czasowych [6]. Przykładem może być jej zastosowanie w analizie szeregów czasowych ARMA(p,q) [23], co jest związane z faktem, że zbiór wszystkich szeregów czasowych ARMA(p,q) ma skończoną parametryzację i w związku z tym tworzy on skończenie wymiarową rozmaitość. Aby dokonać analizy porównawczej dwóch szeregów czasowych biorąc pod uwagę problemy ich aproksymacji, estymacji oraz redukcji wymiaru, analizowanie pojedynczego szeregu czasowego jest niewystarczające i okazuje się koniecznym rozważanie własności całej przestrzeni tych szeregów wraz z ich strukturą geometryczną [6].

2.2.2.1 Transformacja Legendre'a pomiędzy parametryzacjami dualnymi

Niech $\Xi \equiv (\xi^i)_{i=1}^d$ oraz $\Theta \equiv (\theta^i)_{i=1}^d$ są wzajemnie dualnymi bazami na \mathcal{S} , zgodnie z relacją (2.88). Zdefiniujemy współrzędne metryki g ze względu na układ współrzędnych $(\xi^i)_{i=1}^d$ jako:

$$g_{ij}^{\xi} := \langle \partial_{\xi^i}, \partial_{\xi^j} \rangle , \quad i, j = 1, 2, \dots, d \quad , \quad \forall P_{\Xi} \in \mathcal{S} , \quad (2.90)$$

a ze względu na układ współrzędnych $(\theta^j)_{j=1}^d$ jako:

$$g_{ij}^{\theta} := \langle \partial_{\theta^i}, \partial_{\theta^j} \rangle , \quad i, j = 1, 2, \dots, d \quad , \quad \forall P_{\Theta} \in \mathcal{S} . \quad (2.91)$$

Przejdźcie od współrzędnych kontrawariantnych $(\theta^j)_{j=1}^d$ w bazie $\partial_{\theta^i} \equiv \partial/\partial\theta^i$ do kowariantnych $(\theta_j)_{j=1}^d$ w bazie $\partial^{\theta_j} \equiv \partial/\partial\theta_j$ ma postać:

$$\theta_j := \sum_{k=1}^d g_{jk}^{\theta} \theta^k \quad \text{oraz} \quad \partial^{\theta_j} := \sum_{k=1}^d g_{\theta}^{jk} \partial_{\theta^k} , \quad j, k = 1, 2, \dots, d \quad , \quad \forall P_{\Theta} \in \mathcal{S} , \quad (2.92)$$

gdzie g_{θ}^{jk} jest (j, k) -składową macierzy odwrotnej do macierzy informacyjnej (g_{jk}^{θ}) . Ze względu na układ współrzędnych $(\theta_j)_{j=1}^d$, współrzędne metryki g_{θ}^{ij} są równe:

$$g_{\theta}^{ij} = \langle \partial^{\theta_i}, \partial^{\theta_j} \rangle , \quad i, j = 1, 2, \dots, d \quad , \quad \forall P_{\Theta} \in \mathcal{S} . \quad (2.93)$$

Podobnie określamy macierz informacyjną (g_{ξ}^{jk}) jako odwrotną do (g_{jk}^{ξ}) .

Rozważmy transformację układu współrzędnych:

$$\partial^{\theta_j} \equiv \frac{\partial}{\partial\theta_j} = \sum_{i=1}^d \frac{\partial\xi^i}{\partial\theta_j} \frac{\partial}{\partial\xi^i} \equiv \sum_{i=1}^d (\partial^{\theta_j} \xi^i) \partial_{\xi^i} , \quad j = 1, 2, \dots, d \quad , \quad \forall P \in \mathcal{S} \quad (2.94)$$

oraz

$$\partial_{\xi^i} \equiv \frac{\partial}{\partial\xi^i} = \sum_{j=1}^d \frac{\partial\theta_j}{\partial\xi^i} \frac{\partial}{\partial\theta_j} \equiv \sum_{j=1}^d (\partial_{\xi^i} \theta_j) \partial^{\theta_j} = \sum_{j=1}^d (\partial_{\xi^i} \theta^j) \partial_{\theta^j} , \quad i = 1, 2, \dots, d , \quad \forall P \in \mathcal{S} . \quad (2.95)$$

Zatem po skorzystaniu z (2.94)-(2.95) oraz warunku (2.88), można dualne metryki (2.90) oraz (2.93) zapisać następująco:

$$g_{ij}^\xi = \frac{\partial \theta_j}{\partial \xi^i} \quad \text{oraz} \quad g_\theta^{ij} = \frac{\partial \xi^i}{\partial \theta_j}, \quad i, j = 1, 2, \dots, d, \quad \forall P \in \mathcal{S}, \quad (2.96)$$

co oznacza również, że macierze informacyjne, $I_F(\Xi) = (g_{ij}^\xi)$ w bazie $\Xi \equiv (\xi^i)_{i=1}^d$ oraz $I_F(\Theta) = (g_\theta^{ij})$ w bazie¹⁶ dualnej $\Theta \equiv (\theta_i)_{i=1}^d$, są względem siebie odwrotne:

$$I_F(\Xi) = I_F^{-1}(\Theta), \quad \forall P \in \mathcal{S}. \quad (2.97)$$

Rozważmy z kolei funkcję $\psi : \mathcal{S} \rightarrow \mathbb{R}$ oraz następujące cząstkowe równanie różniczkowe:

$$\partial_{\xi^i} \psi = \theta_i, \quad i = 1, 2, \dots, d \quad \text{tzn.} \quad d\psi = \sum_{i=1}^d \theta_i d\xi^i, \quad \forall P_\Xi \in \mathcal{S}. \quad (2.98)$$

Ze względu na (2.96) równanie (2.98) daje:

$$\partial_{\xi^i} \partial_{\xi^j} \psi = g_{ij}^\xi, \quad i, j = 1, 2, \dots, d, \quad \forall P_\Xi \in \mathcal{S}. \quad (2.99)$$

Ze względu na dodatnią określoność metryki g_{ij}^ξ , równanie to oznacza, że druga pochodna ψ tworzy również dodatnio określoną macierz. Zatem ψ jest ściśle wypukłą funkcją współrzędnych $\xi^1, \xi^2, \dots, \xi^d$, dla każdego $P \in \mathcal{S}$.

Podobnie, rozważając funkcję $\phi : \mathcal{S} \rightarrow \mathbb{R}$ oraz cząstkowe równanie różniczkowe:

$$\partial^{\theta_i} \phi = \xi^i, \quad i = 1, 2, \dots, d \quad \text{tzn.} \quad d\phi = \sum_{i=1}^d \xi^i d\theta_i, \quad \forall P_\Theta \in \mathcal{S}. \quad (2.100)$$

Ze względu na (2.96) równanie (2.100) daje:

$$\partial^{\theta_i} \partial^{\theta_j} \phi = g_\theta^{ij}, \quad i, j = 1, 2, \dots, d, \quad \forall P_\Theta \in \mathcal{S}, \quad (2.101)$$

w związku z czym dodatnia określoność dualnej metryki g_θ^{ij} oznacza, że druga pochodna ϕ tworzy dodatnio określoną macierz. Zatem ϕ jest ściśle wypukłą funkcją współrzędnych $\theta^1, \theta^2, \dots, \theta^d$, dla każdego $P \in \mathcal{S}$.

Transformacja Legendre'a: Powiedzmy, że ψ jest pewnym rozwiązaniem równania (2.99). Wtedy po skorzystaniu z (2.96) oraz (2.98), widać, że od $\psi \equiv \psi(\Xi)$ do $\phi \equiv \phi(\Theta)$ można przejść przez *transformację Legendre'a*¹⁷:

$$\phi(\Theta) = \sum_{i=1}^d \xi^i \theta_i - \psi(\Xi), \quad \forall P \in \mathcal{S}. \quad (2.103)$$

¹⁶W zależności od kontekstu i aby nie komplikować zapisu, przez Θ będziemy rozumieli parametr wektorowy we współrzędnych kowariantnych $(\theta_i)_{i=1}^d$ bądź kontrawariantnych $(\theta^i)_{i=1}^d$. Analogicznie postąpimy dla Ξ .

¹⁷Sprawdźmy zgodność (2.103) z warunkami (2.98) oraz (2.100). Z (2.103) otrzymujemy:

$$d\phi = \sum_{i=1}^d \xi^i d\theta_i + \sum_{i=1}^d d\xi^i \theta_i - d\psi, \quad \forall P \in \mathcal{S}. \quad (2.102)$$

Korzystając z (2.98) otrzymujemy $d\phi = \sum_{i=1}^d d\xi^i \theta_i$ jak w (2.100).

Podobnie, powiedzmy, że ϕ jest pewnym rozwiązaniem równania (2.101). Wtedy poprzez transformację Legendre'a:

$$\psi(\Xi) = \sum_{i=1}^d \xi^i \theta_i - \phi(\Theta) , \quad \forall P \in \mathcal{S} , \quad (2.104)$$

można przejść od funkcji ϕ do ψ .

Uwaga: W ogólności transformacje pomiędzy układami współrzędnych $\Xi = (\xi^i)_{i=1}^d$ oraz $\Theta = (\theta_j)_{j=1}^d$, które mają postać (2.103) i (2.103) nazywamy transformacjami Legendre'a¹⁸.

Określenie potencjałów: Funkcje ψ oraz ϕ spełniające odpowiednio warunki (2.98) oraz (2.100), pomiędzy którymi można przejść transformacją Legendre'a (2.103) lub (2.104), nazywamy *potencjałami* układów współrzędnych (odpowiednio Ξ oraz Θ).

Poniżej podamy twierdzenie podsumowujące powyższe rozważania.

Twierdzenie o dualnych układach współrzędnych: Niech $\Xi = (\xi^i)_{i=1}^d$ jest ∇ -afinicznym układem współrzędnych na dualnie płaskiej przestrzeni $(\mathcal{S}, g, \nabla, \nabla^*)$. Wtedy, ze względu na metrykę g , istnieje dualny do $(\xi^i)_{i=1}^d$ układ współrzędnych $\Theta = (\theta_i)_{i=1}^d$, który jest ∇^* -afinicznym układem współrzędnych. Oba te układy współrzędnych są związane transformacją Legendre'a zadaną przy potencjałach $\psi(\Xi)$ oraz $\phi(\Theta)$ poprzez związki (2.103) lub (2.104). Ponadto współrzędne metryki w tych układach współrzędnych są zadane jako drugie pochodne potencjałów, jak w (2.99) oraz (2.101).

Współczynniki koneksji dla układów dualnych: Na koniec podajmy wyprowadzone z użyciem związku (2.82) oraz (2.99) postacie współczynników koneksji afinicznej $\Gamma_{ij,l}^{\xi*}$ [6] (por. (2.55)):

$$\Gamma_{ij,l}^{\xi*} := \langle \nabla_{\partial_{\xi^i}}^* \partial_{\xi^j}, \partial_{\xi^l} \rangle = \partial_{\xi^i} \partial_{\xi^j} \partial_{\xi^l} \psi(\Xi) , \quad i, j, l = 1, 2, \dots, d , \quad \forall P_{\Xi} \in \mathcal{S} \quad (2.107)$$

oraz z użyciem (2.82) oraz (2.101), współczynniki koneksji afinicznej $\Gamma_{\theta}^{ij,l}$:

$$\Gamma_{\theta}^{ij,l} := \langle \nabla_{\partial_{\theta^i}} \partial_{\theta^j}, \partial_{\theta^l} \rangle = \partial_{\theta^i} \partial_{\theta^j} \partial_{\theta^l} \phi(\Theta) , \quad i, j, l = 1, 2, \dots, d , \quad \forall P_{\Theta} \in \mathcal{S} , \quad (2.108)$$

przy czym, ponieważ układy współrzędnych $(\xi^i)_{i=1}^d$ oraz $(\theta_i)_{i=1}^d$ są afiniczne, zatem:

$$\Gamma_{ij,l}^{\xi} = \Gamma_{\theta}^{*ij,l} = 0, \quad i, j, l = 1, 2, \dots, d , \quad \forall P \in \mathcal{S} . \quad (2.109)$$

¹⁸Jeśli ψ oraz ϕ są wypukłymi funkcjami na wypukłych przestrzeniach parametrów V_{Ξ} oraz V_{Θ} , gdzie parametry wektorowe mają postać $\Xi \equiv (\xi^i)_{i=1}^d$ oraz $\Theta \equiv (\theta_i)_{i=1}^d$, to transformacje Legendre'a można sformułować w sposób bardziej ogólny [6]:

$$\phi(\Theta) = \max_{(\Xi \in V_{\Xi})} \left\{ \sum_{i=1}^d \xi^i \theta_i - \psi(\Xi) \right\} , \quad \forall P \in \mathcal{S} \quad (2.105)$$

dla $\psi(\Xi)$ będącego rozwiązaniem równania (2.99). Podobnie, powiedzmy, że $\phi(\Theta)$ jest pewnym rozwiązaniem równania (2.101). Wtedy poprzez transformację Legendre'a:

$$\psi(\Xi) = \max_{(\Theta \in V_{\Theta})} \left\{ \sum_{i=1}^d \xi^i \theta_i - \phi(\Theta) \right\} , \quad \forall P \in \mathcal{S} , \quad (2.106)$$

otrzymujemy potencjał $\psi(\Xi)$.

2.2.3 Geometryczne sformułowanie teorii estymacji dla EFI

Dokładne sformułowanie metody ekstremalnej fizycznej informacji (EFI) jest treścią kolejnych rozdziałów skryptu. Poniżej podajemy jedynie jej wstępną charakterystykę z punktu widzenia geometrii różniczkowej na \mathcal{S} .

Teoria estymacji w metodzie EFI może być określona geometrycznie w sposób następujący. Załóżmy, że z pewnych powodów teoretyczny rozkład P na \mathcal{B} leży na pewnej podprzestrzeni (*warstwie*) $\mathcal{S}_w \subseteq \mathcal{S}$. Warstwa \mathcal{S}_w oraz wymiar \mathcal{B} nie są z góry określone. Zakładamy również, że wszystkie rozważane rozkłady, łącznie z empirycznym, leżą w przestrzeni statystycznej \mathcal{S} , która (w przeciwieństwie do estymacji w statystyce klasycznej [6, 21]) *nie ma znanej postaci metryki* Fishera-Rao. Na podstawie danych, które dały rozkład empiryczny P_{Obs} , szukamy punktu należącego do \mathcal{S}_w , który spełnia zasady informacyjne i jest szukanym oszacowaniem $P_{\hat{\Xi}}$ rozkładu teoretycznego P .

Uwaga o estymacji w statystyce klasycznej: W statystyce klasycznej wyznaczamy krzywą geodezyjną biegnącą przez punkt empiryczny P_{Obs} i szukane oszacowanie $P_{\hat{\Xi}}$ stanu układu leżącego na \mathcal{S}_w . Ponieważ za wyjątkiem przypadku $\alpha = 0$ okazuje się, że α -koneksja nie jest metryczna (tzn. nie jest wyprowadzona jedynie z metryki, w naszym przypadku metryki Fishera-Rao), zatem odległość P_{Obs} od $P_{\hat{\Xi}}$ nie jest na ogół najmniejsza z możliwych [6]. Geodezyjna ta przecina \mathcal{S}_w w pewnym punkcie należącym do \mathcal{S} , który jest szukanym oszacowaniem $P_{\hat{\Xi}}$ stanu układu. Jak wspomnieliśmy, w statystyce klasycznej istnieje jedno ułatwiające estymację założenie. Otóż znana jest ogólna postać modelu statystycznego \mathcal{S} , zatem znana jest i metryka Fishera-Rao na \mathcal{S} .

Uwaga o estymacji w EFI: W przeciwieństwie do tego estymacja w metodzie EFI nie może założyć z góry znajomości postaci metryki g Fishera-Rao. Metoda EFI musi wyestymować g i *estymacja ta jest dynamiczna*, poprzez konstrukcję odpowiednich *zasad informacyjnych*.

Zasada wariacyjna: Jedną z tych zasad powinna zapewnić, że po wyestymowaniu metryki Fishera-Rao, znalezione oszacowanie $P_{\hat{\Xi}}$ metody EFI będzie również leżeć na geodezyjnej łączącej je z P_{Obs} . Stąd pojawia się konieczność wprowadzenia *wariacyjnej zasady informacyjnej*.

Zasada strukturalna: Druga tzw. *strukturalna zasada informacyjna* zapewni, że szukane oszacowanie będzie leżeć w klasie rozwiązań *analitycznych* w parametrze Ξ , w znaczeniu równoważności metrycznej otrzymanego modelu z modelem analitycznym.

Zatem zasada wariacyjna i strukturalna wyznaczają samospójnie punkt $P_{\hat{\Xi}}$ i tym samym wskazują podprzestrzeń statystyczną \mathcal{S}_w . Jednak koneksja afiniczna, dla której układ współrzędnych wzdłuż krzywej geodezyjnej łączącej P_{Obs} z $P_{\hat{\Xi}}$ jest płaski, nie jest α -koneksją Amariego.

Sformułowaniem i zastosowaniem zasad informacyjnych w estymacji metodą EFI zajmiemy się w kolejnych rozdziałach.

2.2.4 Uwaga o rozwinięciu rozkładu w szereg Taylora

W całej treści skryptu zakładamy, że rozkład prawdopodobieństwa $P(\Xi)$ (lub jego logarytm $\ln P(\Xi)$), jest wystarczająco gładki, tzn. posiada *rozwinięcie w szereg Taylora* wystarczająco wysokiego rzędu, w każdym punkcie (pod)przestrzeni statystycznej \mathcal{S} [10]. Zatem albo jest spełniony warunek analityczności rozkładu we wszystkich składowych estymowanego parametru Ξ , albo przynajmniej rozkład prawdopodobieństwa określony w otoczeniu punktu $P \in \mathcal{S}$ posiada dżet $J_P^r(\mathcal{S}, \mathbb{R})$ wystarczająco wysokiego, choć skończonego rzędu r , np. wyrazy do drugiego (lub innego określo-

nego, wyższego) rzędu rozwinięcia w szeregu Taylora, podczas gdy wyższe niż r rzędy rozwinięcia znikają [24]. Do rozważań na temat rzędu dżetów powrócimy w Rozdziale (4.2).

Przestrzeń wektorowa dżetów: Istotną sprawą jest fakt, że zbiór wszystkich r -dżetów funkcji w punkcie $P \in \mathcal{S}$ tworzy skończenie wymiarową przestrzeń wektorową, natomiast ich suma $J^r(\mathcal{S}, \mathbb{R}) = \bigcup_{P \in \mathcal{S}} J_P^r(\mathcal{S}, \mathbb{R})$ jest *wiązką wektorową*, czyli *wiązką włóknistą*¹⁹, której włókno jest przestrzenią wektorową nad przestrzenią bazową \mathcal{S} .

Można pokazać, że również $J_P^\infty(\mathcal{S}, \mathbb{R})$ jest przestrzenią wektorową. Zatem dżety należące do $J_P^\infty(\mathcal{S}, \mathbb{R})$ można dodawać i mnożyć przez liczbę. Tworzą one też algebrę co oznacza, że można je mnożyć. Ważność przestrzeni $J_P^\infty(\mathcal{S}, \mathbb{R})$ ujawnia się przy określeniu rozwinięcia funkcji w szereg Taylora.

Klasy równoważności dżetów: O funkcjach mówimy, że są w tej samej klasie równoważności dżetów, gdy mają takie samo rozwinięcie Taylora.

Pojęcie odwzorowania Taylora na \mathcal{S} : Niech T_P^* jest przestrzenią wektorową dualną do przestrzeni stycznej T_P na przestrzeni statystycznej \mathcal{S} , oraz niech $S^k(T_P^*)$ jest przestrzenią wektorową wszystkich *symetrycznych*²⁰ wieloliniowych odwzorowań:

$$\underbrace{T_P \times \cdots \times T_P}_{k\text{-razy}} \rightarrow \mathbb{R}. \quad (2.111)$$

¹⁹Określenie wiązki włóknistej: (Różniczkowalna) wiązka włóknista $(E, \pi, \mathcal{M}, F, G)$ nad \mathcal{M} składa się z następujących siedmiu elementów [25]:

1. Różniczkowalnej rozmaiłości E nazywanej *przestrzenią totalną*.
2. Różniczkowalnej rozmaiłości \mathcal{M} nazywanej *przestrzenią bazową*.
3. Różniczkowalnej rozmaiłości F nazywanej (*typowym*) *włóknem*.
4. Odwzorowania suriektywnego $\pi : E \rightarrow \mathcal{M}$ nazywanego *rzutowaniem*, którego odwrotny obraz $\pi^{-1}(p) = F_p \cong F$ nazywamy włóknem w p gdzie $p \in \mathcal{M}$.
5. Grupy Liego G nazywanej *grupą strukturalną*, która działa lewostronnie na F .
6. Zbioru otwartych pokryć $\{U_i\}$ rozmaiłości \mathcal{M} z dyfeomorfizmem $\phi_i : U_i \times F \rightarrow \pi^{-1}(U_i)$ takim, że $\pi \circ \phi_i(p, f) = p$, gdzie $f \in F$. Ponieważ ϕ_i^{-1} odwzorowuje $\pi^{-1}(U_i)$ na iloczyn prosty $U_i \times F$ dlatego odwzorowanie ϕ_i jest nazywane *lokalną trywializacją*.
7. Wprowadźmy oznaczenie $\phi_{i,p}(f) \equiv \phi_i(p, f)$. Odwzorowanie $\phi_{i,p} : F \rightarrow F_p$ jest dyfeomorfizmem. Rządamy aby na $U_i \cap U_j \neq \emptyset$ odwzorowanie $t_{ij} \equiv \phi_{i,p}^{-1} \circ \phi_{j,p} : F \rightarrow F$ było elementem grupy G . Wtedy ϕ_i oraz ϕ_j są związane poprzez gładkie odwzorowanie $t_{ij} : U_i \cap U_j \rightarrow G$ w następujący sposób: $\phi_j(p, f) = \phi_i(p, t_{ij}(p)f)$.

Odwzorowania t_{ij} są nazywane *funkcjami przejścia*.

Oznaczenie: Czasami na oznaczenie wiązki włóknistej $(E, \pi, \mathcal{M}, F, G)$ używa się skróconego zapisu $E \xrightarrow{\pi} \mathcal{M}$ lub nawet tylko E .

Przykład: W powyższych rozważaniach przestrzenią bazową \mathcal{M} jest przestrzeń statystyczna \mathcal{S} . Gdy T_P jest przestrzenią styczną do \mathcal{S} w P a T_P^* jest przestrzenią wektorową dualną do T_P , wtedy typowe włókno F_P może być np. przestrzenią tensorową:

$$[T_P]_r^q \equiv \underbrace{T_P \otimes \cdots \otimes T_P}_{q\text{-razy}} \otimes \underbrace{T_P^* \otimes \cdots \otimes T_P^*}_{r\text{-razy}}, \quad (2.110)$$

gdzie q jest indeksem stopnia kontrawariantnego iloczynu tensorowego T_P a r indeksem stopnia kowariantnego iloczynu tensorowego T_P^* . Informacja Fishera jest szczególnym przykładem tensora $[T_P]_2^0$ na \mathcal{S} . Gdy typowe włókno F jest przestrzenią tensorową a przestrzeń wektorowa T_P jest d -wymiarowa, to grupa strukturalna G jest w ogólności ogólną grupą liniowych transformacji $GL(d)$.

²⁰Odwzorowanie nazywamy symetrycznym jeśli jest symetryczne ze względu na permutację zmiennych.

Rozwinięcie w szereg Taylora T pewnej funkcji (np. $P(\Xi)$ lub $\ln P(\Xi)$) na \mathcal{S} określa wtedy odwzorowanie:

$$T : J_P^\infty(\mathcal{S}, \mathbb{R}) \rightarrow S(T_P^*) \quad \text{gdzie} \quad S(T_P^*) \equiv \bigoplus_{k \geq 0} S^k(T_P^*), \quad (2.112)$$

nazywane *odwzorowaniem Taylora*. Szeregi Taylora spełniają istotną rolę w analizie statystycznej [24], o czym przekonamy się przy wyprowadzeniu podstawowego narzędzia estymacji metody EFI, a mianowicie strukturalnej zasady informacyjnej (por. Rozdział 3.1.1).

2.3 Twierdzenie Rao-Cramera i DORC

Estymatory MNW mają asymptotycznie optymalne własności, tzn. są nieobciążone, zgodne, efektywne i dostateczne [11]. Poniższy rozdział poświęcimy efektywności nieobciążonych estymatorów parametru dla dowolnej wielkości próby N .

Estymator efektywny: Wartość dolnego ograniczenia na wariancję estymatora, czyli wariancję estymatora efektywnego, podaje poniższe twierdzenie Rao-Cramera. Jego sednem jest stwierdzenie, że osiągnięcie przez estymator dolnej granicy wariancji podanej w twierdzeniu oznacza, że w klasie estymatorów nieobciążonych, które spełniają warunek regularności (tzn. mają funkcję rozkładu prawdopodobieństwa nie posiadającą punktów nieciągłości zależnych od estymowanego parametru Θ), nie znajdziemy estymatora z mniejszą wariancją.

Estymator efektywny ma więc najmniejszą z możliwych wariancji, jaką możemy uzyskać w procesie estymacji parametru.

2.3.1 Skalarne Twierdzenie Rao-Cramera

Twierdzenie Rao-Cramera (TRC). Przypadek skalarny: Niech $F(\tilde{Y})$ będzie nieobciążonym estymatorem funkcji skalarne parametru $g(\theta)$, tzn.:

$$E_{\theta} F(\tilde{Y}) = g(\theta) \quad (2.113)$$

oraz niech $I_F(\theta)$ będzie informacją Fishera dla parametru θ wyznaczoną na podstawie próby \tilde{Y} . Zakładając warunki regularności, otrzymujemy:

$$\sigma^2_{\theta} F(\tilde{Y}) \geq \frac{[g'(\theta)]^2}{I_F(\theta)}, \quad (2.114)$$

co jest tezą twierdzenia Rao-Cramera. W szczególnym przypadku, gdy $g(\theta) = \theta$, wtedy z (2.114) otrzymujemy następującą postać nierówności Rao-Cramera:

$$\sigma^2_{\theta} F(\tilde{Y}) \geq \frac{1}{I_F(\theta)}. \quad (2.115)$$

Wielkość:

$$\frac{[g'(\theta)]^2}{I_F(\theta)} \quad \text{lub} \quad \frac{1}{I_F(\theta)} \quad \text{dla} \quad g(\theta) = \theta \quad (2.116)$$

nazywana jest dolnym ograniczeniem Rao-Cramera (DORC)²¹. Przypomnijmy, że ponieważ statystyka $F(\tilde{Y})$ jest estymatorem parametru θ , więc wartości jakie przyjmuje nie zależą od tego parametru.

Uwaga: W przypadku, gdy rozkład zmiennej Y traktowany jako funkcja estymowanego parametru θ ma dla pewnych wartości tego parametru punkty nieciągłości, wtedy wariancja estymatora parametru θ występująca po lewej stronie (2.115) może okazać się mniejsza niż wartość po stronie prawej. Sytuacji nieciągłości rozkładu w parametrze nie będziemy jednak rozważali. Przeciwnie, *zakładamy, że rozkład $P(\Theta)$, i jej logarytm $\ln P(\Theta)$, jest wystarczająco gładki*, tzn. posiada rozwinięcie w szereg Taylora wystarczająco wysokiego rzędu, w każdym punkcie (pod)przestrzeni statystycznej \mathcal{S} , jak o tym wspomnieliśmy w Rozdziale 2.2.4.

²¹W jęz. angielskim *Cramér-Rao lower bound* (CRLB).

2.3.1.1 Dowód TRC (wersja dowodu dla przypadku skalarnego)

Współczynnik korelacji liniowej Pearsona dla dwóch zmiennych losowych $S(\tilde{Y})$ i $F(\tilde{Y})$ zdefiniowany jest następująco:

$$\rho_{\theta}(S, F) = \frac{\text{cov}_{\theta}(S, F)}{\sqrt{\sigma_{\theta}^2(S)}\sqrt{\sigma_{\theta}^2(F)}}. \quad (2.117)$$

Z klasycznej analizy statystycznej wiemy, że $\rho_{\theta}(S, F) \in [-1, 1]$, stąd z (2.117) otrzymujemy:

$$\sigma_{\theta}^2(F) \geq \frac{|\text{cov}_{\theta}(S, F)|^2}{\sigma_{\theta}^2(S)}. \quad (2.118)$$

Równość występuje jeżeli współczynnik korelacji liniowej Pearsona jest równy 1, co zachodzi, gdy zmienne S i F są idealnie skorelowane.

Niech teraz zmienna losowa S będzie statystyką wynikową $S(\theta) \equiv S(\tilde{Y}|\theta)$.

Pokażmy, że:

$$g'(\theta) = \text{cov}_{\theta}(S(\theta), F) . \quad (2.119)$$

Istotnie, ponieważ $E_{\theta}(S(\theta)) = 0$, (2.25), zatem:

$$\begin{aligned} \text{cov}_{\theta}(S(\theta), F(\tilde{Y})) &= E_{\theta}(S(\theta)F(\tilde{Y})) = \int_{\mathcal{B}} dy P(y|\theta) S(\theta) F(y) = \int_{\mathcal{B}} dy P(y|\theta) \frac{\partial}{\partial \theta} \frac{P(y|\theta)}{P(y|\theta)} F(y) \\ &= \int_{\mathcal{B}} dy \frac{\partial}{\partial \theta} P(y|\theta) F(y) = \frac{\partial}{\partial \theta} \int_{\mathcal{B}} dy P(y|\theta) F(y) = \frac{\partial}{\partial \theta} E_{\theta} F(\tilde{Y}) = g'(\theta) . \end{aligned} \quad (2.120)$$

Skoro więc zgodnie z (2.28) zachodzi, $\sigma_{\theta}^2 S(\theta) = I_F(\theta)$, więc wstawiając (2.120) do (2.118) otrzymujemy (2.114), co kończy dowód TRC.

2.3.1.2 Przykład skalarny DORC dla rozkładu normalnego

Interesuje nas *przypadek estymacji skalarnego parametru* $\theta = \mu$ w próbie prostej \tilde{Y} , przy czym zakładamy, że $g(\mu) = \mu$. Rozważmy średnią arytmetyczną $\bar{Y} = \frac{1}{N} \sum_{n=1}^N Y_n$ (z realizacją $\bar{y} = \frac{1}{N} \sum_{n=1}^N y_n$), która, zakładając jedynie identyczne rozkłady zmiennych Y_i próby, jest dla dowolnego rozkładu $p(\mathbf{y})$ zmiennej Y , nieobciążonym estymatorem wartości oczekiwanej $\mu \equiv E_{\mu}(Y) = \int_{\mathcal{Y}} dy p(\mathbf{y}) y$, tzn.:

$$E_{\mu}(\bar{Y}) = \int_{\mathcal{B}} dy \bar{y} P(y|\mu) = E_{\mu}(Y) = \mu . \quad (2.121)$$

Ponadto dla próby prostej, z bezpośredniego rachunku otrzymujemy:

$$\sigma_{\mu}^2(\bar{Y}) = \int_{\mathcal{B}} dy P(y|\mu) (\bar{y} - E(\bar{Y}))^2 = \frac{\sigma^2}{N} , \quad (2.122)$$

gdzie σ^2 jest wariancją $\sigma_{\mu}^2(Y)$ zmiennej Y .

Niech teraz zmienna pierwotna Y ma rozkład normalny $N(\mu, \sigma^2)$. Ze związków (1.17) oraz (1.18)-(1.19) wiemy, że średnia \bar{Y} jest estymatorem MNW parametru $\mu = E(Y)$, zatem przyjmijmy $F(\tilde{Y}) = \hat{\mu} = \bar{Y}$. Z (2.122) widzimy więc, że dla zmiennych o rozkładzie normalnym zachodzi:

$$\sigma_{\mu}^2(F) = \sigma_{\mu}^2(\bar{Y}) = \frac{\sigma^2}{N} . \quad (2.123)$$

W przypadku rozkładu normalnego warunek (2.115) łatwo sprawdzić bezpośrednim rachunkiem. Istotnie, korzystając z (2.9) oraz (2.17), otrzymujemy (por. (2.17)):

$$\sigma_\mu^2 S(\mu) = \sigma_\mu^2 \left(\frac{N}{\sigma^2} (\bar{Y} - \mu) \right) = \left(\frac{N}{\sigma^2} \right)^2 \frac{\sigma^2}{N} = \frac{N}{\sigma^2} = I_F(\mu). \quad (2.124)$$

Z (2.123) oraz (2.124) otrzymujemy:

$$\sigma_\mu^2(\hat{\mu}) = \frac{1}{I_F(\mu)}, \quad (2.125)$$

co stanowi DORC (2.116) dla nierówności Rao-Cramera (2.115). Warunek ten otrzymaliśmy już poprzednio dla rozkładu normalnego (por. (2.18)). Spełnienie go oznacza, że średnia arytmetyczna \bar{Y} jest efektywnym²² estymatorem wartości oczekiwanej zmiennej losowej opisanej rozkładem normalnym $N(\mu, \sigma^2)$.

Uwaga o rozkładach eksponentialnych: Rozkład normalny jest szczególnym przypadkiem szerszej klasy rozkładów, które spełniają warunek DORC. Rozkłady te są tzw. rozkładami eksponentialnymi (2.61) wprowadzonymi w Rozdziale 2.2. Powyżej, w przypadku rozkładu normalnego, sprawdziliśmy ten fakt bezpośrednim rachunkiem zakładając w pierw typ rozkładu zmiennej Y , a potem sprawdzając, że estymator $\hat{\mu}$ parametru $\mu = E_\mu(Y)$ osiąga DORC.

2.3.2 Wieloparametrowe Twierdzenie Rao-Cramera

Gdy dokonujemy równoczesnej estymacji $d > 1$ parametrów, wtedy funkcja wynikowa $S(\Theta)$ jest d -wymiarowym wektorem kolumnowym (1.9), natomiast obserwowana IF w punkcie $\hat{\Theta}$, czyli $iF(\hat{\Theta})$, oraz wartość oczekiwana z $iF(\Theta)$, czyli I_F (por.(2.14)), są $d \times d$ wymiarowymi macierzami.

Analogia inflacji wariancji: Poniżej pokażemy, że włączenie do analizy dodatkowych parametrów ma (na ogół) wpływ na wartość IF dla interesującego nas, wyróżnionego parametru. Sytuacja ta jest analogiczna do problemu inflacji wariancji estymatora parametru w analizie częstotliwościowej [14]. Poniżej przedstawiona zostanie odnosząca się do tego problemu wieloparametrowa wersja twierdzenia o dolnym ograniczeniu w nierówności Rao-Cramera (DORC).

Uwaga o wersjach TRC: Poniżej podamy dwie równoważne wersje [6] wieloparametrowego Twierdzenia Rao-Cramera (TRC). Pierwsza z nich okaże się być bardzo użyteczna przy wprowadzeniu w Rozdziale 2.7 relacji pomiędzy tzw. informacją Stam'a a pojemnością informacyjną układu. W Rozdziale 2.8.1 przekonamy się, że wersja druga TRC jest użyteczna w sprawdzeniu czy wieloparametrowa estymacja przebiega na DORC.

²²Tzn. pośród estymatorów nieobciążonych parametru $\mu = E_\mu(Y)$ i regularnych, posiada najmniejszą z możliwych wariancji.

2.3.2.1 Pierwsza wersja wieloparametrowego TRC

Wieloparametrowe Twierdzenie RC (wersja pierwsza): Niech $F(\tilde{Y})$ będzie funkcją skalarną z wartością oczekiwaną:

$$E_{\Theta} F(\tilde{Y}) = g(\Theta) \in \mathbf{R} \quad (2.126)$$

oraz $I_F(\Theta)$ niech będzie oczekiwaną informacją Fishera (2.14) dla $\Theta \equiv (\theta_i)_{i=1}^d$ wyznaczoną na przestrzeni próby \mathcal{B} . Zachodzi wtedy nierówność:

$$\sigma_{\Theta}^2(F(\tilde{Y})) \geq \mathbf{a}^T I_F^{-1}(\Theta) \mathbf{a}, \quad (2.127)$$

gdzie I_F^{-1} jest macierzą odwrotną do macierzy informacyjnej Fishera I_F , natomiast

$$\mathbf{a} = \frac{\partial g(\Theta)}{\partial \Theta} \quad (2.128)$$

jest d -wymiarowym wektorem.

Uwaga: Warunek (2.126) oznacza, że skalarna funkcja $F(\tilde{Y})$ jest nieobciążonym estymatorem skalarnej funkcji $g(\Theta)$ wektorowego parametru Θ .

2.3.2.2 Przykład wektorowego DORC

Jako ilustrację powyższego wieloparametrowego Twierdzenia RC przedstawimy przykład, przyjmując szczególną postać skalarnej funkcji $g(\Theta)$, o której zakładamy, że jest liniową funkcją składowych θ_i wektora parametrów [13]:

$$g(\Theta) = \mathbf{a}^T \Theta = \sum_{i=1}^d a_i \theta_i, \quad (2.129)$$

gdzie \mathbf{a} jest pewnym znanym wektorem o stałych składowych a_i , które nie zależą od składowych wektora Θ . Załóżmy chwilowo, że $\mathbf{a}^T = (1, 0, \dots, 0)$, tzn. jedynie $a_1 \neq 0$. Wtedy z (2.129) otrzymujemy $g(\Theta) = \theta_1$, natomiast (2.127) w Twierdzeniu RC, $\sigma_{\Theta}^2(F) \geq \mathbf{a}^T I_F^{-1}(\Theta) \mathbf{a}$, przyjmuje dla rozważanego nieobciążonego estymatora F parametru θ_1 , tzn. $E_{\Theta}(F(\tilde{Y})) = \theta_1$, postać:

$$\sigma_{\Theta}^2(F) \geq [I_F^{-1}(\Theta)]_{11} =: I_F^{11}(\Theta), \quad (2.130)$$

gdzie $I_F^{11}(\Theta)$ oznacza element (1,1) macierzy $I_F^{-1}(\Theta)$. Prawa strona nierówności (2.130) podaje dolne ograniczenie wariancji estymatora F , pod warunkiem, że θ_1 jest wyróżnionym parametrem a wartości pozostałych parametrów *nie są znane*. Oznaczmy wewnętrzną strukturę $d \times d$ -wymiarowych macierzy $I_F(\Theta)$ oraz $I_F(\Theta)^{-1}$ następująco:

$$I_F(\Theta) = \begin{pmatrix} I_{F11} & I_{F12} \\ I_{F21} & I_{F22} \end{pmatrix} \quad (2.131)$$

oraz

$$I_F^{-1}(\Theta) = \begin{pmatrix} I_F^{11} & I_F^{12} \\ I_F^{21} & I_F^{22} \end{pmatrix} \quad (2.132)$$

gdzie I_{F11} oraz $I_F^{11} = [I_F^{-1}(\Theta)]_{11}$ (zgodnie z oznaczeniem wprowadzonym w (2.130)) są liczbami, I_{F22} , I_F^{22} są $(d-1) \times (d-1)$ -wymiarowymi macierzami, natomiast $(I_{F12})_{1 \times (d-1)}$, $(I_{F21})_{(d-1) \times 1}$,

$(I_F^{12})_{1 \times (d-1)}$, $(I_F^{21})_{(d-1) \times 1}$ odpowiednimi wierszowymi bądź kolumnowymi wektorami o wymiarze $(d-1)$.

Rozważmy parametr θ_1 . Jego informacja Fishera (patrz poniżej **Uwaga o nazwie**) jest równa $I_{F11} = I_{F11}(\theta_1)$. *Nie oznacza to jednak, że $\sigma_\Theta^2(F)$ oraz I_{F11} są z sobą automatycznie powiązane nierównością $\sigma_\Theta^2(F) \geq 1/I_{F11}$, która jest treścią Twierdzenia RC (2.115).* Udowodniliśmy ją bowiem tylko dla przypadku parametru skalarne, tzn. gdy tylko jeden parametr jest estymowany, a reszta parametrów jest znana.

Określmy relację pomiędzy $(I_{F11})^{-1}$ oraz I_F^{11} . Oczywiście zachodzi:

$$I_F(\Theta) \cdot I_F^{-1}(\Theta) = \begin{pmatrix} I_{F11} & I_{F12} \\ I_{F21} & I_{F22} \end{pmatrix} \begin{pmatrix} I_F^{11} & I_F^{12} \\ I_F^{21} & I_F^{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad (2.133)$$

gdzie $\mathbf{1}$ jest $(d-1) \times (d-1)$ -wymiarową macierzą jednostkową.

Z powyższego mamy:

$$\begin{aligned} (I_{F11})_{1 \times 1} (I_F^{11})_{1 \times 1} + (I_{F12})_{1 \times (d-1)} (I_F^{21})_{(d-1) \times 1} &= 1 \\ \Rightarrow (I_F^{11})^{-1} &= I_{F11} + I_{F12} I_F^{21} (I_F^{11})^{-1}, \end{aligned} \quad (2.134)$$

$$\begin{aligned} (I_{F21})_{(d-1) \times 1} (I_F^{11})_{1 \times 1} + (I_{F22})_{(d-1) \times (d-1)} (I_F^{21})_{(d-1) \times 1} &= (0)_{(d-1) \times 1} \\ \Rightarrow I_F^{21} &= -(I_{F22})^{-1} I_{F21} I_F^{11} \end{aligned} \quad (2.135)$$

skąd otrzymujemy:

$$(I_F^{11})^{-1} = I_{F11} - I_{F12} (I_{F22})^{-1} I_{F21}. \quad (2.136)$$

Ponieważ I_{F22} jest macierzą informacyjną (dla parametrów $\theta_2, \theta_3, \dots, \theta_d$), jest więc ona zgodnie z rozważaniami przedstawionymi poniżej (2.32), symetryczna i nieujemnie określona. Symetryczna i nieujemnie określona jest zatem $(I_{F22})^{-1}$. Ponieważ z symetrii macierzy I_F wynika $I_{F12} = (I_{F21})^T$, zatem ostatecznie forma kwadratowa $I_{F12} (I_{F22})^{-1} I_{F21} \geq 0$, stąd z (2.136) otrzymujemy:

$$(I_F^{11})^{-1} \leq I_{F11} \Rightarrow I_F^{11} \geq \frac{1}{I_{F11}}, \quad (2.137)$$

co zgodnie z (2.130) oznacza, że:

$$\sigma_\Theta^2(F) \geq I_F^{11} \geq \frac{1}{I_{F11}}. \quad (2.138)$$

Wniosek: Zatem widzimy, że I_F^{11} daje *silniejsze ograniczenie* niż $(I_{F11})^{-1}$. Tzn. w przypadku estymacji wieloparametrowej należy zastosować związek $\sigma^2(F) \geq I_F^{11}$, (2.130), gdyż to właśnie on jest właściwy na podstawie wieloparametrowego Twierdzenia RC. Zastosowanie $\sigma_\Theta^2(F) \geq 1/I_{F11}$, tak jak byśmy mieli do czynienia z przypadkiem skalarnym, może błędnie zaniżyć wartość dolnego ograniczenia na $\sigma_\Theta^2(F)$.

Uwaga o nazwie I_{F11} : W “statycznie” ukierunkowanej analizie statystycznej wielkość $(I_F^{11})^{-1}$ jest interpretowana jako informacja Fishera dla θ_1 – jednak w treści skryptu odstawimy od tej nazwy. Okazuje się, że w analizie ukierunkowanej na estymację “dynamiczną”, tzn. generującą równania różniczkowe dla rozkładów, bardziej użyteczne jest nazwać I_F^{11} po prostu *dolnym ograniczeniem RC* na

wariancję estymatora parametru θ_1 w sytuacji gdy pozostałe parametry są nieznane (tzn. trzeba je estymować z próby równocześnie z θ_1). Natomiast I_{F11} będziemy nazywali, zgodnie z tym jak to uczyniliśmy, informacją Fishera parametru θ_1 i to niezależnie od tego czy inne parametry są równocześnie estymowane, czy też nie.

Podsumowanie na temat zaniżenia DORC: Należy pamiętać, że estymując parametr θ_1 należy być świadomym faktu występowania równoczesnej estymacji innych parametrów, gdyż wstawienie wartości I_{F11} do nierówności RC może w przypadku estymacji wieloparametrowej doprowadzić do zaniżenia wartości dolnego ograniczenia wariancji tego parametru.

Przypadek “pseudo-skalarny”: Istnieje jednak pewien wyjątek spowodowany dokładnym zerowaniem się I_{F12} dla dowolnego N . Wtedy z (2.136) wynika, że wzrost wariancji estymatora parametru związany z dodaniem nowych parametrów o nieznanych wartościach byłby równy zeru. Tak też było w rozważanym wcześniej przykładzie rozkładu normalnego $N(\mu, \sigma)$ (porównaj (2.24) z (2.17)).

Podobnie, taki szczególny przypadek zachodzi, gdy wieloparametrowym rozkładem prawdopodobieństwa jest wiarygodność N -wymiarowej próby $P(\theta_1, \theta_2, \dots, \theta_N) = \prod_{n=1}^N p_{\theta_n}$, gdzie każdy estymowany parametr θ_n określa tylko jeden punktowy rozkład p_{θ_n} . Wtedy macierz informacyjna Fishera I_F jest diagonalna i zachodzi $I_{Fnn} = (I_F^{nn})^{-1}$, a w miejsce (2.138) otrzymujemy dla każdego parametru θ_n :

$$\sigma_{\Theta}^2(F_n) \geq I_F^{nn} = \frac{1}{I_{Fnn}}, \quad n = 1, 2, \dots, N, \quad \text{gdy } I_F \text{ diagonalne}, \quad (2.139)$$

jako szczególny przypadek nierówności RC, gdzie F_n jest estymatorem $\hat{\theta}_n$ parametru θ_n .

2.3.2.3 Druga wersja wieloparametrowego TRC

Niech $\hat{\Theta} \equiv \hat{\Theta}(\tilde{Y}) = (\hat{\theta}_i(\tilde{Y}))_{i=1}^d$ jest nieobciążonym estymatorem parametru $\Theta = (\theta_i)_{i=1}^d$, co oznacza, że zachodzi:

$$\Theta = E_{\Theta} [\hat{\Theta}(\tilde{Y})], \quad \forall P_{\Theta} \in \mathcal{S}. \quad (2.140)$$

(Oczekiwana) macierz kowariancji²³ $V_{\Theta}[\hat{\Theta}]$ nieobciążonego estymatora $\hat{\Theta}$ w bazie Θ nazywamy $d \times d$ -wymiarową macierz o elementach:

$$V_{\Theta ij}[\hat{\Theta}] := E_{\Theta} [(\hat{\theta}_i(\tilde{Y}) - \theta_i)(\hat{\theta}_j(\tilde{Y}) - \theta_j)], \quad i, j = 1, 2, \dots, d, \quad \forall P_{\Theta} \in \mathcal{S}. \quad (2.141)$$

Zachodzi następujące twierdzenie.

Wieloparametrowe Twierdzenie RC (wersja druga): Macierz kowariancji $V_{\Theta}[\hat{\Theta}]$ nieobciążonego estymatora $\hat{\Theta}$ spełnia nierówność [6]:

$$V_{\Theta}[\hat{\Theta}] \geq I_F^{-1}(\Theta) \quad \forall P_{\Theta} \in \mathcal{S}, \quad (2.142)$$

co oznacza, że macierz $V_{\Theta}[\hat{\Theta}] - I_F^{-1}(\Theta)$ jest dodatnio półokreślona.

Estymator efektywny: Nieobciążony estymator $\hat{\Theta}$ spełniający równość w nierówności (2.142):

$$V_{\Theta}[\hat{\Theta}] = I_F^{-1}(\Theta) \quad \forall P_{\Theta} \in \mathcal{S}. \quad (2.143)$$

nazywamy estymatorem *efektywnym* parametru Θ .

²³Czyli tzw. macierzą oczekiwanego błędu kwadratowego.

2.4 Entropia informacyjna Shannona i entropia względna

W rozdziale tym omówimy pojęcie, które podaje *globalną charakterystykę* pojedynczego rozkładu prawdopodobieństwa, tzn. entropię Shannona. Dokładniejsze omówienie własności entropii Shannona można znaleźć w [20]. Z treści poniższych przykładów wynika jaki jest rozmiar próby N .

Entropia Shannona: Niech $P(\omega)$ będzie rozkładem prawdopodobieństwa²⁴ określonym na przestrzeni zdarzeń Ω , gdzie ω jest punktem w Ω . Jeśli przestrzeń zdarzeń Ω jest dyskretna, to *informacyjna entropia Shannona* jest zdefiniowana następująco:

$$S_H(P) = -k \sum_{\omega \in \Omega} P(\omega) \ln P(\omega), \quad (2.144)$$

gdzie k jest liczbą dodatnią²⁵, którą przyjmujemy dalej jako równą 1.

Rozkład prawdopodobieństwa jako element sympleksu rozkładów: Niech Ω jest rozpięta przez skończoną liczbę \aleph elementów będących możliwymi wynikami doświadczenia, w rezultacie którego otrzymujemy wartości y_i , $i = 1, 2, \dots, \aleph$, zmiennej losowej Y . Rozkład prawdopodobieństwa P jest wtedy reprezentowany przez wektor $\vec{p} = (p_i)_{i=1}^{\aleph}$, należący do *sympleksu rozkładów prawdopodobieństwa* [20], tzn. jego \aleph składowych spełnia warunki:

$$p_i \geq 0 \quad \text{oraz} \quad \sum_{i=1}^{\aleph} p_i = 1. \quad (2.145)$$

m-Sympleks rozkładów prawdopodobieństwa, czyli zbiór $\Delta^m = \{(p_1, p_2, \dots, p_m, p_{m+1}) \in \mathbf{R}^{m+1}; p_j \geq 0, j = 1, \dots, m+1 \text{ gdzie } m+1 \leq \aleph, \text{ dla } \sum_{j=1}^{m+1} p_j \leq 1\}$ jest *zbiorem wypukłym*, tzn. każdy punkt $\vec{p}_m \in \Delta^m$ można przedstawić jako $\vec{p}_m = \sum_{i=1}^{m+1} \lambda_i p_i$, gdzie $\sum_{i=1}^{m+1} \lambda_i = 1$.

Entropia Shannona (2.144) rozkładu (2.145) ma postać:

$$S_H(P) = - \sum_{i=1}^{\aleph} p_i \ln p_i. \quad (2.146)$$

Zapis $S_H(P)$, gdzie w argumencie pominięto oznaczenie zmiennej losowej Y podkreśla, że *jedyną rozważaną przez nas cechą zmiennej losowej jest jej rozkład prawdopodobieństwa P* .

²⁴**Przestrzeń stanów modelu:** Mówimy, że na przestrzeni zdarzeń Ω została określona funkcja $\omega \rightarrow P(\omega)$ spełniająca warunki, $P(\omega) \geq 0$ oraz $\sum_{\omega} P(\omega) = 1$, nazywana wtedy miarą probabilistyczną. Zbiór wszystkich miar probabilistycznych określonych na Ω tworzy *przestrzeń stanów modelu*.

²⁵W przypadku statystycznej entropii fizycznej \aleph może być np. *liczbą konfiguracji określonej liczby molekuł przy zadanej energii całkowitej układu*. Wtedy k jest utożsamiane ze stałą Boltzmann'a k_B . Dla układu określonego w przestrzeni ciągłej \mathbb{R}^3 liczba konfiguracji jest nieskończona. Gdyby ograniczyć się do skończonej podprzestrzeni i podzielić ją na komórki o skończonej wielkości, i podobnie uczynić w przestrzeni pędowej, to liczba możliwych konfiguracji układu byłaby skończona a jego entropia mogłaby być policzona. Jednakże poprawny rachunek entropii wymaga wtedy utożsamienia konfiguracji powiedzmy n cząstek różniących się jedynie ich permutacją, w ramach jednej klasy równoważności. Na fakt, że właściwa przestrzeń próby ma w tym przypadku nie \aleph lecz $\aleph/n!$ punktów zwrócił uwagę Gibbs, a otrzymaną przestrzeń próby nazywa się *przestrzenią próby Gibbsa*. Np. dla 1 cm^3 cieczy w zwykłych warunkach $n \approx 10^{23}$. Problem ten nie będzie rozważany dalej w niniejszym skrypcie.

2.4.1 Interpretacja entropii Shannona

Maksymalna możliwa wartość entropii Shannona²⁶ wynosi $\ln \aleph$ i jest osiągnięta, gdy wszystkie wyniki są równo prawdopodobne ($p_i = 1/\aleph$), tzn. gdy stan układu jest maksymalnie zmieszany.

Układ w stanie czystym: Gdy jeden z wyników jest pewny, wtedy tylko jedna, odpowiadająca mu współrzędna wektora \vec{p} jest równa jeden, a pozostałe są równe 0. Mówimy wtedy, że układ znajduje się w stanie czystym, a odpowiadająca mu wartość entropii Shannona jest minimalna i równa zero.

Entropia jako miara niepewności wyniku eksperymentu: Z powyższych przykładów można wnioskować, że entropię Shannona można interpretować jako *miarę niepewności otrzymania wyniku eksperymentu* będącego realizacją rozkładu prawdopodobieństwa P lub inaczej, jako wielkość informacji koniecznej *do określenia* wyniku, który może się pojawić w rezultacie przeprowadzenia eksperymentu na układzie.

Podstawowe własności entropii Shannona można znaleźć w [20].

2.4.2 Przypadek ciągłego rozkładu prawdopodobieństwa

Przejdźcie z dyskretnego do ciągłego rozkładu prawdopodobieństwa polega na zastąpieniu sumowania w (2.146) całkowaniem po całym zakresie zmienności zmiennej losowej Y . W ten sposób otrzymujemy Boltzmanowską postać entropii Shannona:

$$S_H(P) = - \int_{-\infty}^{+\infty} dy P(y) \ln P(y). \quad (2.147)$$

Jednakże dla pewnych funkcji rozkładu $P(y)$, całka (2.147) może nie być określona.

Czysty stan klasyczny: Jako ilustrację powyższego stwierdzenia rozważmy sytuację [20], gdy rozkład $P(y)$ przyjmuje w przedziale $[0, t]$ wartość t^{-1} i zero wszędzie poza tym przedziałem. Wtedy jego entropia $S_H(P)$ jest równa $\ln t$ i dla $t \rightarrow 0$ dąży do *minus nieskończoności*. Procedura ta odpowiada przejściu do punktowego, *czystego stanu klasycznego* opisanego dystrybucją delta Diraca, dla której $S_H(P) = -\infty$.

Zatem przyjmując poziom zerowy entropii jako punkt odniesienia, dokładne określenie stanu opisanego deltą Diraca (której fizycznie mógłby odpowiadać nieskończony skok w gęstości rozkładu substancji cząstki) wymaga dostarczenia nieskończonej ilości informacji o układzie. Do sprawy powrócimy w jednym z kolejnych rozdziałów.

Problem transformacyjny $S_H(P)$: Definicja (2.147) ma pewien formalny minus, związany z brakiem porządknych własności transformacyjnych entropii Shannona. Omówimy go poniżej.

Ze względu na unormowanie prawdopodobieństwa do jedności, gęstość rozkładu prawdopodobieństwa przekształca się przy transformacji układu współrzędnych tak jak odwrotność objętości.

²⁶Ze względu na warunek $\sum_{i=1}^{\aleph} p_i = 1$ maksymalizujemy funkcję $S_{H\text{ war}}(P) = S_H(P) - \lambda \left(1 - \sum_{i=1}^{\aleph} p_i\right)$, licząc pochodną po p_j , $j = 1, 2, \dots, \aleph$, gdzie λ jest czynnikiem Lagrange'a.

Przykład zmiennej jednowymiarowej: Z unormowania $\int_{-\infty}^{+\infty} dy P(y) = 1$ wynika, że $P(y)$ musi transformować się tak, jak $1/y$.

Rozważmy transformację układu współrzędnych $y \rightarrow y'$. Różniczka zmiennej Y transformuje się wtedy zgodnie z $dy' = J(\frac{y}{y'})dy$, gdzie J jest jacobianem transformacji, natomiast rozkład prawdopodobieństwa transformuje się następująco: $P'(y') = J^{-1}(\frac{y}{y'})P(y)$. Zatem tak jak to powinno być, unormowanie rozkładu w transformacji pozostaje niezmiennicze, tzn.:

$$\int_{-\infty}^{+\infty} dy P(y) = \int_{-\infty}^{+\infty} dy' P'(y') = 1. \quad (2.148)$$

Rozważmy teraz entropię Shannona układu określoną dla rozkładu ciągłego jak w (2.147), $S_H(P) = - \int dy P(y) \ln P(y)$. Jak to zauważyliśmy powyżej, entropia układu jest miarą nieuporządkowania w układzie, bądź informacji potrzebnej do określenia wyniku eksperymentu. Zatem również i ona powinna być niezmiennicza przy rozważanej transformacji. Niestety, chociaż miara probabilistyczna pozostaje niezmiennicza, to ponieważ $\ln P(y) \neq \ln P'(y')$ zatem $S_H(P) \neq S_H(P')$.

Tak więc *logarytm z gęstości rozkładu prawdopodobieństwa nie jest niezmienniczy przy transformacji układu współrzędnych* i w konsekwencji otrzymujemy następujący wniosek, słuszny również w przypadku rozkładu prawdopodobieństwa wielowymiarowej zmiennej losowej Y .

Wniosek: Entropia Shannona (2.147) nie jest niezmiennicza przy transformacji układu współrzędnych przestrzeni bazowej \mathcal{B} .

Z drugiej strony, ze względu na wyjątkowe pośród innych entropii własności entropii Shannona dla rozkładu dyskretnego [20], zrezygnowanie z jej ciągłej granicy (2.147) mogłoby się okazać decyzją chybioną. Również jej związek z informacją Fishera omówiony dalej, przekonuje o istotności pojęcia entropii Shannona w jej formie ciągłej.

Entropia względna jako rozwiązanie problemu transformacji: Proste rozwiązanie zaistniałego problemu polega na zaobserwowaniu, że ponieważ iloraz dwóch gęstości $P(y)$ oraz $P_{ref}(y)$ transformuje się jak skalar, tzn.:

$$\frac{P'(y')}{P'_{ref}(y')} = \frac{J^{-1}(\frac{y}{y'})P(y)}{J^{-1}(\frac{y}{y'})P_{ref}(y)} = \frac{P(y)}{P_{ref}(y)}, \quad (2.149)$$

gdzie $P_{ref}(y)$ występuje jako pewien rozkład referencyjny, zatem wielkość nazywana *entropią względną*:

$$S_H(P|P_{ref}) \equiv \int_{-\infty}^{+\infty} dy P(y) \ln \frac{P(y)}{P_{ref}(y)}, \quad (2.150)$$

posiada już własność niezmienniczości:

$$S_H(P'|P'_{ref}) = S_H(P|P_{ref}) \quad (2.151)$$

przy transformacji układu współrzędnych.

Entropia względna jako wartość oczekiwana: Zwróćmy uwagę, że entropia względna jest wartością oczekiwaną logarytmu dwóch rozkładów, $P(\mathbf{y})$ oraz $P_{ref}(\mathbf{y})$, zmiennej Y :

$$S_H(P|P_{ref}) = E_P \left(\ln \frac{P(Y)}{P_{ref}(Y)} \right), \quad (2.152)$$

wyznaczoną przy założeniu, że zmienna losowa Y ma rozkład P .

Entropia względna a analiza doboru modelu: W ten sposób problem logarytmu ilorazu funkcji wiarygodności (czy w szczególności dewiancji) wykorzystywanego w analizie braku dopasowania modelu (por. Rozdziały 1.2-1.3), powrócił w postaci konieczności wprowadzenia entropii względnej. Istotnie, wiemy, że pojęcie logarytmu ilorazu rozkładów okazało się już użyteczne w porównywaniu modeli statystycznych i wyborze modelu bardziej “wiarygodnego”. Wybór modelu powinien być niezmienny ze względu na transformację układu współrzędnych przestrzeni bazowej \mathcal{B} . Logarytm ilorazu rozkładów posiada żadaną własność. Jego wartością oczekiwaną, która jest entropią względną, zajmujemy się w Rozdziale (2.4.3). Przez wzgląd na związek IF z entropią względną dla rozkładów różniących się infitezymalnie mało, jej pojęcie będzie nam towarzyszyło do końca skryptu.

Nazwy entropii względnej: Entropię względną (2.150) nazywana się również entropią Kullbacka-Leiblera (KL) lub dywergencją informacji.

2.4.3 Entropia względna jako miara odległości

Rozważmy eksperyment, którego wyniki są generowane z pewnego określonego, chociaż nieznanego rozkładu prawdopodobieństwa P_{ref} należącego do obszaru \mathcal{O} . Obszar \mathcal{O} jest nie posiadającym izolowanych punktów zbiorem rozkładów prawdopodobieństwa, który jest przestrzenią metryczną zupełną. Oznacza to, że na \mathcal{O} można określić odległość oraz każdy ciąg Cauchy’ego ma granicę należącą do \mathcal{O} .

Twierdzenie Sanova [26]: Jeśli mamy N - wymiarową próbkę niezależnych pomiarów pochodzących z rozkładu prawdopodobieństwa P_{ref} pewnej zmiennej losowej, to prawdopodobieństwo Pr , że empiryczny rozkład (częstości) \hat{P} wypadnie w obszar \mathcal{O} , spełnia asymptotycznie związek:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \ln Pr \left\{ \hat{P} \in \mathcal{O} \right\} = -\beta \quad \text{gdzie} \quad \beta = \inf_{P_{\mathcal{O}} \in \mathcal{O}} S_H(P_{\mathcal{O}}|P_{ref}), \quad (2.153)$$

który w przybliżeniu można zapisać następująco:

$$Pr(\hat{P} \in \mathcal{O}) \sim e^{-NS_H(P_{\mathcal{O}}|P_{ref})}, \quad (2.154)$$

gdzie $P_{\mathcal{O}}$ jest rozkładem należącym do \mathcal{O} różnym od P_{ref} z najmniejszą wartością entropii względnej $S_H(P_{\mathcal{O}}|P_{ref})$. Rozkład $P_{\mathcal{O}}$ uznajemy za rozkład wyestymowany na podstawie empirycznego rozkładu częstości \hat{P} otrzymanego w obserwacji.

Wniosek: Zauważmy, że Twierdzenie Sanova jest rodzajem prawa wielkich liczb, zgodnie z którym dla wielkości próby N dążącej do nieskończoności, *prawdopodobieństwo zaobserwowania rozkładu częstości \hat{P} należącego do \mathcal{O} różnego od prawdziwego rozkładu P_{ref} (tzn. tego który generował wyniki eksperymentu), dąży do zera.*

Fakt ten wyraża właśnie relacja (2.154), a ponieważ $S_H(P_{\mathcal{O}}|P_{ref})$ jest w jej wykładniku, zatem entropia względna określa tempo w jakim prawdopodobieństwo $Pr(\hat{\mathcal{P}})$ dąży do zera wraz ze wzrostem N .

Entropia względna dla rozkładów dyskretnych: Jeśli $P_{ref} \equiv (p^i)_{i=1}^{\aleph}$ jest dyskretnym rozkładem prawdopodobieństwa o \aleph możliwych wynikach, wtedy rozkład $P_{\mathcal{O}} \equiv (p_{\mathcal{O}}^i)_{i=1}^{\aleph} \in \mathcal{O}$ jest też dyskretnym rozkładem o \aleph wynikach, a entropia względna $S_H(P_{\mathcal{O}}|P_{ref})$ ma postać:

$$S_H(P_{\mathcal{O}}|P_{ref}) = \sum_{i=1}^{\aleph} p_{\mathcal{O}}^i \ln \frac{p_{\mathcal{O}}^i}{p^i}. \quad (2.155)$$

Dla rozkładów ciągłych entropia względna została określona w (2.150). Przekonamy się, że entropia względna jest miarą określającą jak bardzo dwa rozkłady różnią się od siebie.

Przykład: W celu ilustracji twierdzenie Sanova założmy, że przeprowadzamy doświadczenie rzutu niesymetryczną monetą z wynikami orzeł, reszka, zatem $\aleph = 2$. Rozkład teoretyczny P_{ref} jest więc zero-jedynkowy. Natomiast w wyniku pobrania N -elementowej próbki dokonujemy jego estymacji na podstawie rozkładu empirycznego $\hat{\mathcal{P}}$ częstości pojawienia się wyników orzeł lub reszka. Zatem:

$$P_{ref} \equiv (p^i)_{i=1}^{\aleph=2} = (p, 1-p) \quad \text{oraz} \quad \hat{\mathcal{P}} \equiv (\hat{p}^i)_{i=1}^{\aleph=2} = \left(\frac{m}{N}, 1 - \frac{m}{N}\right). \quad (2.156)$$

Twierdzenie Bernoulliego mówi, że prawdopodobieństwo pojawienia się wyniku orzeł z częstością m/N w N -losowaniach wynosi:

$$Pr(\hat{\mathcal{P}}) \equiv Pr\left(\frac{m}{N}\right) = \binom{N}{m} p^m (1-p)^{N-m}. \quad (2.157)$$

Biorąc logarytm naturalny obu stron (2.157), następnie stosując słuszne dla dużego n przybliżenie Stirlinga, $\ln n! \approx n \ln n - n$, dla każdej silni w wyrażeniu $\binom{N}{m}$, i w końcu biorąc eksponentę obu stron, można otrzymać [20]:

$$Pr(\hat{\mathcal{P}}) \approx e^{-NS_H(\hat{\mathcal{P}}|P_{ref})}, \quad (2.158)$$

gdzie

$$\begin{aligned} S_H(\hat{\mathcal{P}}|P_{ref}) &= \left[\frac{m}{N} \left(\ln \frac{m}{N} - \ln p \right) + \left(1 - \frac{m}{N} \right) \left(\ln \left(1 - \frac{m}{N} \right) - \ln(1-p) \right) \right] \\ &= \sum_{i=1}^{\aleph=2} \hat{p}^i \ln \frac{\hat{p}^i}{p^i}. \end{aligned} \quad (2.159)$$

W ostatniej równości skorzystano z (2.156) otrzymując entropię względną (2.155) dla przypadku liczby wyników $\aleph = 2$.

Podsumowanie: W powyższym przykładzie entropia względna $S_H(\hat{\mathcal{P}}|P_{ref})$ pojawiła się jako pojęcie wtórne, wynikające z wyznaczenia prawdopodobieństwa otrzymania w eksperymencie empirycznego rozkładu częstości $\hat{\mathcal{P}} \in \mathcal{O}$ jako oszacowania rozkładu P_{ref} . Fakt ten oznacza, że entropia względna nie jest tworem sztucznym, wprowadzonym do teorii jedynie dla wygody jako “jakaś” miara odległości pomiędzy rozkładami, lecz, że jest właściwą dla przestrzeni statystycznej rozkładów

miarą probabilistyczną tej odległości, tzn. “dywergencją informacji” pomiędzy rozkładami. Okaże się też, że spośród innych miar odległości jest ona wyróżniona poprzez jej związek ze znaną już nam informacją Fishera.

W końcu podamy twierdzenie o dodatniości entropii względnej, które jeszcze bardziej przybliży nas do zrozumienia entropii jako miary odległości pomiędzy rozkładami, rozwijanego w treści następnego rozdziału.

Twierdzenie (Nierówność informacyjna): Jeśli $P(\mathbf{y})$ oraz $P_{ref}(\mathbf{y})$ są dwoma rozkładami gęstości prawdopodobieństwa, wtedy entropia względna spełnia następującą nierówność:

$$S_H(P(\mathbf{y})|P_{ref}(\mathbf{y})) \geq 0, \quad (2.160)$$

i nierówność ta jest ostra za wyjątkiem przypadku, gdy $P(\mathbf{y}) = P_{ref}(\mathbf{y})$ [13].

2.5 Geometria przestrzeni rozkładów prawdopodobieństwa i metryka Rao-Fishera

Pojęcie metryki Fishera wprowadziliśmy już w Rozdziale 2.2. Do ujęcia tam przedstawionego dojdziemy jeszcze raz, wychodząc od pojęcia entropii względnej.

Rozważmy dwa rozkłady prawdopodobieństwa: $P = (p^i)_{i=1}^N$ oraz $P' = P + dP = (p'^i = p^i + dp^i)_{i=1}^N$, różniące się *infinitesimalnie mało*, przy czym $p_i \neq 0$ dla każdego i . Rozkłady P oraz P' spełniają warunek unormowania $\sum_{i=1}^N p^i = 1$ oraz $\sum_{i=1}^N p^i + dp^i = 1$, skąd:

$$\sum_{i=1}^N dp^i = 0. \quad (2.161)$$

Ponieważ dp^i/p^i jest wielkością infinitesimalnie małą, zatem z rozwinięcia:

$$\ln(1 + \frac{dp^i}{p^i}) = \frac{dp^i}{p^i} - 1/2(\frac{dp^i}{p^i})^2 + \dots \quad (2.162)$$

otrzymujemy, że entropia KL rozkładów P oraz P' wynosi:

$$S_H(P|P + dP) = \sum_{i=1}^N p^i \ln \frac{p^i}{p^i + dp^i} = - \sum_{i=1}^N p^i \ln \frac{p^i + dp^i}{p^i} \approx \frac{1}{2} \sum_{i=1}^N \frac{dp^i dp^i}{p^i}. \quad (2.163)$$

Ostatnia postać $S_H(P|P + dP)$ sugeruje, że entropia KL określa w naturalny sposób na przestrzeni statystycznej \mathcal{S} infinitesimalny kwadrat odległości pomiędzy rozważanymi rozkładami, co oznaczałoby również *lokalne* określenie na przestrzeni statystycznej \mathcal{S} pewnej metryki. Do jej związku z metryką Rao-Fisher’a wprowadzoną w Rozdziale 2.2 powrócimy później.

Infinitesimalny interwał w \mathcal{S} : Niech $d\vec{p} = (dp^1, \dots, dp^N)$ jest infinitesimalnym wektorem w przestrzeni rozkładów prawdopodobieństwa, spełniającym warunek (2.161). Wprowadzając na przestrzeni statystycznej \mathcal{S} aparat matematyczny geometrii różniczkowej [6], zapiszmy kwadrat różniczkowego interwału w tej przestrzeni następująco:

$$ds^2 = \sum_{i,j=1}^N g_{ij} dp^i dp^j. \quad (2.164)$$

W celu uzgodnienia (2.163) z (2.164) wprowadźmy na \mathcal{S} metrykę:

$$g_{ij} = \frac{\delta_{ij}}{p^i} . \quad (2.165)$$

Związek (2.164) oznacza, że *liczba możliwych wyników* \aleph określa wymiar przestrzeni \mathcal{S} oraz, że entropia względna KL definiuje dla infinitezymalnie bliskich rozkładów symetryczną metrykę na przestrzeni statystycznej \mathcal{S} , pozwalającą mierzyć odległość pomiędzy tymi rozkładami. Zastępując prawdopodobieństwa p^i częstościami, metryka ta staje się statystyką związaną z obserwowaną informacją Fishera wprowadzoną poprzednio.

Powyżej stan układu określony był w reprezentacji rozkładu prawdopodobieństwa zmiennej losowej, tzn. określony był przez podanie rozkładu prawdopodobieństwa. Dwóm infinitezymalnie blisko leżącym stanom opowiadały rozkłady P oraz P' .

Reprezentacja amplitudowa \mathcal{S} : Gdy interesują nas czysto geometryczne własności metryki Fishera, wtedy wygodne jest użycie innej reprezentacji do opisu stanu układu, określonej następująco. Niech $Q \equiv (q^i)_{i=1}^{\aleph}$ oraz $Q' \equiv (q'^i)_{i=1}^{\aleph} = (q^i + dq^i)_{i=1}^{\aleph}$ opisują te same co poprzednio, dwa infinitezymalnie blisko leżące stany układu tyle, że zarządzamy, aby infinitezymalny kwadrat interwału pomiędzy nimi (2.164) był równy:

$$ds^2 = 4 \sum_{i=1}^{\aleph} dq^i dq^i . \quad (2.166)$$

Porównując formułę (2.166) z (2.164) zauważamy, że zgadzają się one z sobą o ile $Q = (q^i)_1^{\aleph}$ oraz $P = (p^i)_1^{\aleph}$ są powiązane związkiem:

$$dq^i = \frac{dp^i}{2\sqrt{p^i}} \quad i = 1, \dots, \aleph , \quad (2.167)$$

co zachodzi wtedy gdy

$$q^i = \sqrt{p^i} \quad i = 1, \dots, \aleph . \quad (2.168)$$

Wielkości q^i nazywamy *amplitudami układu*²⁷. Definiują one na \mathcal{S} nowe współrzędne, dla których $q^i \geq 0$ dla każdego i .

\mathcal{S} z geometrią jednostkowej sfery: Otrzymana w bazie q^i geometria \mathcal{S} jest geometrią jednostkowej sfery, tzn. ze względu na unormowanie rozkładu prawdopodobieństwa do jedności, amplitudy spełniają następujący warunek unormowania na promieniu jednostkowym:

$$\sum_{i=1}^{\aleph} p^i = \sum_{i=1}^{\aleph} q^i q^i = 1 . \quad (2.169)$$

Na sferze tej możemy określić odległość geodezyjną D_{Bhatt} , tzw. odległość Bhattacharyya' pomiędzy dwoma rozkładami prawdopodobieństwa P oraz P' , jako długość kątową liczoną wzdłuż koła wielkiego pomiędzy dwoma wektorami Q oraz Q' o składowych będących amplitudami $q^i = \sqrt{p^i}$ oraz $q'^i = \sqrt{p'^i}$.

²⁷ Fisher korzystał z amplitud prawdopodobieństwa niezależnie od ich pojawienia się w mechanice kwantowej.

Odległości Bhattacharyya: Kwadrat infinitezimalnego interwału (2.166) jest przykładem odległości Bhattacharyya, którą ogólnie określamy następująco: Jeśli $P = (p^i)$ oraz $P' = (p'^i)$ są rozkładami prawdopodobieństwa, wtedy odległość Bhattacharyya pomiędzy Q oraz Q' jest iloczynem wewnętrznym określonym na przestrzeni statystycznej \mathcal{S} następująco:

$$\cos D_{Bhatt} = \sum_{i=1}^N q^i q'^i = \sum_{i=1}^N \sqrt{p^i p'^i} \equiv B(p, p') . \quad (2.170)$$

Statystyczna odległością Bhattacharyya wygląda więc jak iloczyn wewnętrzny mechaniki kwantowej. W kolejnych rozdziałach przekonamy się, że nie jest to błędne skojarzenie.

Hessian entropii Shannona: Zauważmy, że metryka g_{ij} jest Hessianem (tzn. macierzą drugich pochodnych) entropii Shannona:

$$g_{ij} = -\partial_i \partial_j S_H(p) = \frac{\partial}{\partial p^i} \frac{\partial}{\partial p^j} \sum_{k=1}^N p^k \ln p^k = \frac{\delta^{ij}}{p^j} \geq 0 , \quad (2.171)$$

zgodnie z (2.165). Powyższy związek oznacza, że fakt wklęsłości entropii Shannona [20] daje dodatnią określoność metryki g_{ij} na \mathcal{S} .

Metryka indukowana z g_{ij} : Uzasadnijmy fakt nazwania czasami metryki g_{ij} , (2.165), metryką Rao-Fishera.

Założmy, że interesuje nas pewna podprzestrzeń przestrzeni rozkładów prawdopodobieństwa \mathcal{S} . Wprowadźmy na niej układ współrzędnych θ^a . Korzystając z (2.165) widać, że metryka $g_{ij} = \frac{\delta^{ij}}{p^j}$ indukuje w tej podprzestrzeni \mathcal{S} metrykę:

$$\begin{aligned} g_{ab} &= \sum_{i,j=1}^N \frac{\partial p^i}{\partial \theta^a} \frac{\partial p^j}{\partial \theta^b} g_{ij} = \sum_{i=1}^N \frac{\partial_a p^i \partial_b p^i}{p^i} = \sum_{i=1}^N p^i \partial_a \ln p^i \partial_b \ln p^i \\ &\equiv E(\partial_a \ell_\Theta \partial_b \ell_\Theta) = E_\Theta \mathbf{iF}(\Theta) = I_F(\Theta) . \end{aligned} \quad (2.172)$$

gdzie skorzystano z zapisu $\ell_\Theta \equiv l(y|\Theta) = \ln P(\Theta) = (\ln p^i(\Theta))_{i=1}^N$ oraz $\partial_a \equiv \partial/\partial \theta^a$, (2.35).

W Rozdziale 2.2 we wzorze (2.36) zdefiniowaliśmy metrykę Rao-Fishera jako $g_{ab} = E_\Theta(\partial_a \ell_\Theta \partial_b \ell_\Theta)$. Tak więc ostatecznie otrzymaliśmy zgodność nazwania metryki (2.165) metryką Rao-Fishera. *Metryka g_{ij} jest jednak wielkością obserwowaną a nie oczekiwaną, tak jak metryka Rao-Fishera g_{ab} .*

Metryka Rao-Fishera zapisana w amplitudach: Korzystając ze związku $q^i = \sqrt{p^i}$, (2.168), pomiędzy prawdopodobieństwami i amplitudami, metrykę Rao-Fishera (2.172) w reprezentacji amplitudowej można zapisać następująco:

$$g_{ab} = 4 \sum_{i=1}^N \frac{\partial q^i}{\partial \theta^a} \frac{\partial q^i}{\partial \theta^b} . \quad (2.173)$$

Przejście do zmiennej ciągłej: Gdy liczba parametrów θ^a (więc i wektorów bazowych $\vec{\theta}^a$) rozpinających osie układu współrzędnych rozważanej podprzestrzeni statystycznej jest skończona, wtedy można dokonać następującego uogólnienia metryki na przypadek ciągłego rozkładu prawdopodobieństwa $P(y)$ zmiennej losowej Y :

$$g_{ab} = \int_{\mathcal{Y}} d\mathbf{y} \frac{\partial_a P}{P} \frac{\partial_b P}{P} = \int_{\mathcal{Y}} d\mathbf{y} P \partial_a \ell_{\Theta} \partial_b \ell_{\Theta} \equiv E(\partial_a \ell_{\Theta} \partial_b \ell_{\Theta}) = E_{\Theta} \mathbf{iF}(\Theta) = I_F(\Theta) , \quad (2.174)$$

gdzie, ponieważ przestrzeń zdarzeń ciągłego rozkładu prawdopodobieństwa jest nieskończenie wymiarowa, w miejsce sumowania po $i = 1, \dots, \aleph$ pojawiło się całkowanie po wartościach $\mathbf{y} \in \mathcal{Y}$ zmiennej losowej Y . Otrzymana postać metryki jest jawnie niezmiennicza ze względu na zmianę układu współrzędnych $\mathbf{y} \rightarrow \mathbf{y}'$ w przestrzeni bazowej \mathcal{Y} .

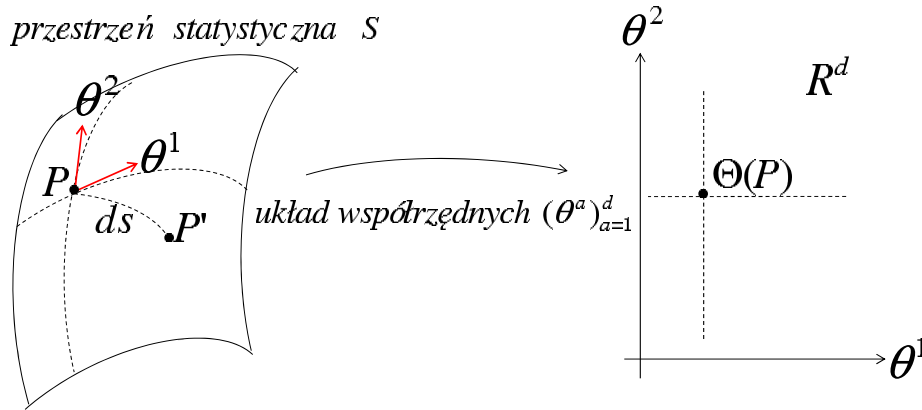
Na koniec, dokonując reparametryzacji i przechodząc do amplitud $q(\mathbf{y}|\Theta) = \sqrt{P}(\mathbf{y}|\Theta)$, można metrykę (2.174) dla przypadku rozkładu ciągłego zapisać następująco:

$$g_{ab} = 4 \int_{\mathcal{Y}} d\mathbf{y} \partial_a q(\mathbf{y}|\Theta) \partial_b q(\mathbf{y}|\Theta) , \quad (2.175)$$

gdzie jawnie zaznaczono zależność amplitudy rozkładu od parametru Θ .

Przypadek funkcji wiarygodności: Powyższa analiza dotyczyła przypadku próby $N = 1$ - wymiarowej. Gdyby rozkładem prawdopodobieństwa P była funkcja wiarygodności próby, wtedy w miejscu zmiennej losowej Y pojawiłaby się próba $\tilde{Y} \equiv (Y_n)_{n=1}^N$, a w miejscu przestrzeni \mathcal{Y} , przestrzeń próby \mathcal{B} . Poza tym, rozważania w obecnym rozdziale pozostałyby takie same [6].

Omówiną sytuację odległości na przestrzeni statystycznej przedstawia graficznie poniższy rysunek dla liczby parametrów $d = 2$.



Dla dwóch nieskończenie bliskich rozkładów:

P oraz $P' = P + dP$

$$ds^2 = \sum_{a,b=1}^d g_{ab} d\theta^a d\theta^b \quad \text{gdzie } g_{ab} \text{ jest metryką Rao-Fishera}$$

Przykład: Wyznaczyć kwadrat infinitezimalnego interwału dla dwóch stanów posiadających rozkład normalny $N(\mu, \sigma)$. Rozkład normalny ma postać:

$$P(\mathbf{y}, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\mathbf{y}-\mu)^2}{2\sigma^2}}, \quad (2.176)$$

gdzie wektor parametrów $\Theta \equiv (\theta^a)_{a=1}^2 = (\mu, \sigma)^T$. Macierz informacyjna Fishera dla rozkładu normalnego $N(\mu, \sigma)$ ma wyznaczoną poprzednio postać (2.24). Zatem metryka Rao-Fishera na 2-wymiarowej przestrzeni normalnych rozkładów prawdopodobieństwa z układem współrzędnych μ, σ ma (dla próby $N = 1$), postać:

$$(g_{ab}) = I_F(\Theta) = E_{\Theta} \mathbf{iF}(\Theta) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}, \quad (2.177)$$

tzn. składowe metryki (2.174) są równe:

$$g_{\mu\mu} = \frac{1}{\sigma^2}, \quad g_{\mu\sigma} = g_{\sigma\mu} = 0, \quad g_{\sigma\sigma} = \frac{2}{\sigma^2}. \quad (2.178)$$

Zatem otrzymany kwadrat infinitezimalnego interwału na 2-wymiarowej (pod)przestrzeni statystycznej \mathcal{S} wynosi [20]:

$$ds^2 = g_{\mu\mu}d\mu^2 + g_{\mu\sigma}d\mu d\sigma + g_{\sigma\sigma}d\sigma^2 = \frac{1}{\sigma^2} (d\mu^2 + 2d\sigma^2). \quad (2.179)$$

Odpowiada on metrze Poincarégo ze stałą ujemną krzywizną. W końcu, z postaci rozkładu (2.176) otrzymujemy również bazę w przestrzeni stycznej do \mathcal{S} :

$$\partial_{\mu} \ln P(\mathbf{y} | (\mu, \sigma)) = \frac{\mathbf{y} - \mu}{\sigma^2}, \quad \partial_{\sigma} \ln P(\mathbf{y} | (\mu, \sigma)) = \frac{(\mathbf{y} - \mu)^2}{\sigma^3} - \frac{1}{\sigma}. \quad (2.180)$$

Wniosek z przykładu: Na osi μ , na której $\sigma = 0$ leżą punkty odpowiadające klasycznym stanom czystym i są one zgodnie z (2.179) nieskończenie daleko odległe od dowolnego punktu we wnętrzu górnej półpłaszczyzny $\sigma > 0$. W konsekwencji oznacza to, że wynik pewny, któremu odpowiada rozkład klasyczny z $\sigma = 0$, jest łatwy do odróżnienia od każdego innego.

2.6 Informacja Fishera

2.6.1 Informacja Fishera jako entropia

W powyższych rozważaniach informacja Fishera I_F pojawiła się poprzez nierówność Rao-Cramera (2.115), jako wielkość określająca graniczną dobroć procedury estymacyjnej parametru rozkładu, tzn. o ile estymator efektywny istnieje, to informacja Fishera określa minimalną wartość jego wariancji. Ponieważ im informacja Fishera mniejsza, tym to graniczne oszacowanie parametru gorsze, zatem jest ona również miarą stopnia nieuporządkowania układu określonego rozkładem prawdopodobieństwa. Takie zrozumienie informacji Fishera odnosi się do typu analizowanego rozkładu prawdopodobieństwa i jako miara nieuporządkowania rozkładu oznacza brak przewidywalności procedury estymacyjnej (spróbuj pomyśleć o sensie oszacowania wartości oczekiwanej rozkładu jednorodnego).

Poniżej uzasadnimy stwierdzenie, że informacja Fishera okazuje się być proporcjonalna do entropii Kullbacka-Leiblera rozkładów różniących się infinitezymalnie mało w parametrze rozkładu [8].

Szukany związek informacji Fishera z entropią Kullbacka-Leiblera pokażemy w trzech krokach. Rozważmy N -wymiarową próbę Y_1, Y_2, \dots, Y_N , gdzie każda ze zmiennych losowych Y_n jest określona na przestrzeni \mathcal{Y} i posiada rozkład prawdopodobieństwa $p(\mathbf{y}_n|\theta)$. Przyjmijmy, dla uproszczenia rozważań, że $\Theta = \theta$ jest parametrem skalarnym.

(1) Funkcja wiarygodności $P(y|\theta)$ ma postać:

$$P(y|\theta) = \prod_{n=1}^N p_n(\mathbf{y}_n|\theta), \quad (2.181)$$

gdzie $y \equiv (\mathbf{y}_n)_{n=1}^N$ jest realizacją próby.

Zauważmy, że informację Fishera parametru θ określoną w (2.14) i (2.3):

$$I_F = - \int_{\mathcal{B}} dy P(y|\theta) \frac{\partial^2 \ln P(y|\theta)}{\partial \theta^2}, \quad (2.182)$$

gdzie \mathcal{B} jest przestrzenią próby, a $dy \equiv d^N \mathbf{y} = d\mathbf{y}_1 d\mathbf{y}_2 \dots d\mathbf{y}_N$, można zapisać następująco:

$$I_F = \int_{\mathcal{B}} dy P(y|\theta) \left(\frac{\partial \ln P(y|\theta)}{\partial \theta} \right)^2. \quad (2.183)$$

Istotnie, postać (2.183) otrzymujemy z (2.182) po skorzystaniu z:

$$\frac{\partial}{\partial \theta} \left(\frac{\partial \ln P}{\partial \theta} \right) = \frac{\partial}{\partial \theta} \left(\frac{1}{P} \frac{\partial P}{\partial \theta} \right) = -\frac{1}{P^2} \left(\frac{\partial P}{\partial \theta} \right)^2 + \frac{1}{P} \left(\frac{\partial^2 P}{\partial \theta^2} \right) \quad (2.184)$$

oraz warunku regularności pozwalającego na wyłączenie różniczkowania po parametrze przed całkę (por. Rozdział 2.1.1):

$$\int_{\mathcal{B}} dy \left(\frac{\partial^2 P}{\partial \theta^2} \right) = \frac{\partial^2}{\partial \theta^2} \int_{\mathcal{B}} dy P = \frac{\partial^2}{\partial \theta^2} 1 = 0. \quad (2.185)$$

(2) **Zadanie:** Pokazać, że zachodzi następujący *rozkład informacji Fishera* parametru θ :

$$I_F = \sum_{n=1}^N I_{Fn} \quad \text{gdzie} \quad I_{Fn} = \int_{\mathcal{Y}} d\mathbf{y}_n p_n(\mathbf{y}_n|\theta) \left(\frac{\partial \ln p_n(\mathbf{y}_n|\theta)}{\partial \theta} \right)^2, \quad (2.186)$$

na składowe informacje I_{Fn} , zależne jedynie od rozkładów punktowych $p_n(\mathbf{y}|\theta)$.

Rozwiązanie: Po skorzystaniu z (2.181) zauważamy, że:

$$\ln P(y|\theta) = \sum_{n=1}^N \ln p_n, \quad \text{zatem} \quad \frac{\partial \ln P}{\partial \theta} = \sum_{n=1}^N \frac{1}{p_n} \frac{\partial p_n}{\partial \theta} \quad (2.187)$$

i podnosząc ostatnie wyrażenie do kwadratu otrzymujemy:

$$\left(\frac{\partial \ln P}{\partial \theta} \right)^2 = \sum_{\substack{n,m=1 \\ m \neq n}}^N \frac{1}{p_m} \frac{1}{p_n} \frac{\partial p_m}{\partial \theta} \frac{\partial p_n}{\partial \theta} + \sum_{n=1}^N \frac{1}{p_n^2} \left(\frac{\partial p_n}{\partial \theta} \right)^2. \quad (2.188)$$

Wstawiając (2.181) oraz (2.188) do (2.183) otrzymujemy:

$$I_F = \int_{\mathcal{B}} dy \prod_{k=1}^N p_k \sum_{\substack{n,m=1 \\ m \neq n}}^N \frac{1}{p_m} \frac{1}{p_n} \frac{\partial p_m}{\partial \theta} \frac{\partial p_n}{\partial \theta} + \int_{\mathcal{B}} dy \prod_{k=1}^N p_k \sum_{n=1}^N \frac{1}{p_n^2} \left(\frac{\partial p_n}{\partial \theta} \right)^2. \quad (2.189)$$

Ze względu na warunek normalizacji rozkładów brzegowych:

$$\int_{\mathcal{Y}} d\mathbf{y}_n p_n(\mathbf{y}_n|\theta) = 1, \quad (2.190)$$

w (2.189) z iloczynu $\prod_{k=1}^N p_k$ pozostaje w pierwszym składniku jedynie $p_m p_n$ (dla $k \neq m$ oraz $k \neq n$), natomiast w drugim składniku pozostaje jedynie p_n (dla $k \neq n$), tzn.:

$$I_F = \sum_{\substack{n,m=1 \\ m \neq n}}^N \int_{\mathcal{Y}} \int_{\mathcal{Y}} d\mathbf{y}_m d\mathbf{y}_n \frac{\partial p_m}{\partial \theta} \frac{\partial p_n}{\partial \theta} + \sum_{n=1}^N \int_{\mathcal{Y}} d\mathbf{y}_n \frac{1}{p_n} \left(\frac{\partial p_n}{\partial \theta} \right)^2. \quad (2.191)$$

W końcu, ze względu na:

$$\int_{\mathcal{Y}} d\mathbf{y}_n \frac{\partial p_n}{\partial \theta} = \frac{\partial}{\partial \theta} \int_{\mathcal{Y}} d\mathbf{y}_n p_n = \frac{\partial}{\partial \theta} 1 = 0, \quad (2.192)$$

pierwszy składnik w (2.191) zeruje się i pozostaje jedynie drugi, zatem:

$$I_F = \sum_{n=1}^N \int_{\mathcal{Y}} d\mathbf{y}_n \frac{1}{p_n} \left(\frac{\partial p_n}{\partial \theta} \right)^2 = \sum_{n=1}^N \int_{\mathcal{Y}} d\mathbf{y}_n p_n \left(\frac{\partial \ln p_n}{\partial \theta} \right)^2. \quad (2.193)$$

Ostatecznie otrzymujemy więc szukany rozkład informacji Fishera (2.186):

$$I_F = \sum_{n=1}^N I_{Fn} \quad \text{gdzie} \quad I_{Fn} = \int_{\mathcal{Y}} d\mathbf{y}_n p_n(\mathbf{y}_n|\theta) \left(\frac{\partial \ln p_n(\mathbf{y}_n|\theta)}{\partial \theta} \right)^2.$$

(3) Zadanie: Znaleźć związek informacji Fishera z entropią Kullbacka-Leiblera rozkładów $P(y|\theta)$ oraz $P(y|\theta + \Delta\theta)$ różniących się infinitesimalnie mało w parametrze θ .

Rozwiązanie: Zastąpmy całkę w (2.186) dla I_{Fn} sumą Riemanna:

$$\begin{aligned} I_{Fn} &= \sum_k \Delta \mathbf{y}_{nk} \frac{1}{p_n(\mathbf{y}_{nk}|\theta)} \left[\frac{p_n(\mathbf{y}_{nk}|\theta + \Delta\theta) - p_n(\mathbf{y}_{nk}|\theta)}{\Delta\theta} \right]^2 \\ &= (\Delta\theta)^{-2} \Delta \mathbf{y}_n \sum_k p_n(\mathbf{y}_{nk}|\theta) \left[\frac{p_n(\mathbf{y}_{nk}|\theta + \Delta\theta)}{p_n(\mathbf{y}_{nk}|\theta)} - 1 \right]^2. \end{aligned} \quad (2.194)$$

Powyższa zamiana całkowania na sumę jest wprowadzona dla wygody i jest ścisłym przejściem w granicy $\Delta \mathbf{y}_{nk} \xrightarrow{k \rightarrow \infty} 0$. W drugiej równości w (2.194) przyjęto równe przyrosty $\Delta \mathbf{y}_{nk} = \Delta \mathbf{y}_n$ dla każdego k , co w tej granicy nie zmienia wyniku. Natomiast przejście dla pochodnej po θ pod całką I_{Fn} w (2.186) dokonane w (2.194) jest słuszne w granicy $\Delta\theta \rightarrow 0$.

W granicy $\Delta\theta \rightarrow 0$ każde z powyższych wyrażen $p_n(\mathbf{y}_{nk}|\theta + \Delta\theta)/p_n(\mathbf{y}_{nk}|\theta)$ dąży do 1. Wtedy każda z wielkości:

$$\delta_{\Delta\theta}^n \equiv \frac{p_n(\mathbf{y}_{nk}|\theta + \Delta\theta)}{p_n(\mathbf{y}_{nk}|\theta)} - 1 \quad (2.195)$$

staje się mała i rozwijając funkcję logarytmu do wyrazu drugiego rzędu:

$$\ln(1 + \delta_{\Delta\theta}^n) \approx \delta_{\Delta\theta}^n - (\delta_{\Delta\theta}^n)^2/2, \quad (2.196)$$

otrzymujemy:

$$(\delta_{\Delta\theta}^n)^2 \approx 2[\delta_{\Delta\theta}^n - \ln(1 + \delta_{\Delta\theta}^n)]. \quad (2.197)$$

Korzystając z (2.195) oraz (2.197) wyrażenie (2.194) można zapisać następująco:

$$\begin{aligned} I_{Fn} &= -2(\Delta\theta)^{-2} \Delta\mathbf{y}_n \left[\sum_k p_n(\mathbf{y}_{nk}|\theta) \ln \left(\frac{p_n(\mathbf{y}_{nk}|\theta + \Delta\theta)}{p_n(\mathbf{y}_{nk}|\theta)} \right) \right. \\ &\quad \left. - \sum_k p_n(\mathbf{y}_{nk}|\theta + \Delta\theta) + \sum_k p_n(\mathbf{y}_{nk}|\theta) \right], \quad \text{dla } \Delta\theta \rightarrow 0. \end{aligned} \quad (2.198)$$

Ze względu na warunek normalizacji zachodzi $\sum_k p_n(\mathbf{y}_{nk}|\theta + \Delta\theta) = \sum_k p_n(\mathbf{y}_{nk}|\theta) = 1$, zatem dwie ostatnie sumy po k w nawiasie kwadratowym znoszą się wzajemnie i (2.198) redukuje się do postaci:

$$\begin{aligned} I_{Fn} &= -2(\Delta\theta)^{-2} \sum_k \Delta\mathbf{y}_n p_n(\mathbf{y}_{nk}|\theta) \ln \left(\frac{p_n(\mathbf{y}_{nk}|\theta + \Delta\theta)}{p_n(\mathbf{y}_{nk}|\theta)} \right) \\ &= -2(\Delta\theta)^{-2} \int_{\mathcal{Y}} d\mathbf{y}_n p_n(\mathbf{y}_n|\theta) \ln \left(\frac{p_n(\mathbf{y}_n|\theta + \Delta\theta)}{p_n(\mathbf{y}_n|\theta)} \right) \\ &= 2(\Delta\theta)^{-2} S_H[p_n(\mathbf{y}_n|\theta) | p_n(\mathbf{y}_n|\theta + \Delta\theta)], \quad \text{dla } \Delta\theta \rightarrow 0, \end{aligned} \quad (2.199)$$

gdzie w ostatnim przejściu skorzystaliśmy z (2.150). Postać I_{Fn} w drugiej linii w (2.199) można również, wykorzystując unormowanie rozkładów $p_m(\mathbf{y}_m|\theta)$, zapisać następująco:

$$I_{Fn} = -2(\Delta\theta)^{-2} \prod_{\substack{m=1 \\ m \neq n}}^N \int_{\mathcal{Y}} d\mathbf{y}_m p_m(\mathbf{y}_m|\theta) \int_{\mathcal{Y}} d\mathbf{y}_n p_n(\mathbf{y}_n|\theta) \ln \left(\frac{p_n(\mathbf{y}_n|\theta + \Delta\theta)}{p_n(\mathbf{y}_n|\theta)} \right), \quad \text{dla } \Delta\theta \rightarrow 0, \quad (2.200)$$

skąd po skorzystaniu z $I_F = \sum_{n=1}^N I_{Fn}$, (2.186), otrzymujemy:

$$\begin{aligned} I_F = \sum_{n=1}^N I_{Fn} &= -2(\Delta\theta)^{-2} \sum_{n=1}^N \prod_{\substack{m=1 \\ m \neq n}}^N \int_{\mathcal{Y}} d\mathbf{y}_m p_m(\mathbf{y}_m|\theta) \int_{\mathcal{Y}} d\mathbf{y}_n p_n(\mathbf{y}_n|\theta) \ln \left(\frac{p_n(\mathbf{y}_n|\theta + \Delta\theta)}{p_n(\mathbf{y}_n|\theta)} \right) \\ &= -2(\Delta\theta)^{-2} \int_{\mathcal{B}} d\mathbf{y} \prod_{m=1}^N p_m(\mathbf{y}_m|\theta) \ln \left(\frac{\prod_{n=1}^N p_n(\mathbf{y}_n|\theta + \Delta\theta)}{\prod_{n=1}^N p_n(\mathbf{y}_n|\theta)} \right) \\ &= 2(\Delta\theta)^{-2} S_H[P(\mathbf{y}|\theta) | P(\mathbf{y}|\theta + \Delta\theta)], \quad \text{dla } \Delta\theta \rightarrow 0. \end{aligned} \quad (2.201)$$

Tak więc (2.201) daje wsparcie dla intuicji wspomnianej na początku obecnego Rozdziału, która mówi, że skoro entropia jest miarą nieuporządkowania układu to informacja Fishera jest również miarą jego nieuporządkowania.

Uwaga o lokalności związku IF z KL: O własnościach globalnych rozkładu wypowiada się entropia Shannona, natomiast wniosek wynikający z (2.201) ma sens tylko dla entropii względnej. Zatem związek ten może być co najwyżej sygnałem, że niektóre własności układu związane z entropią Shannona mogą być ujęte w języku informacji Fishera. Na niektóre sytuacje, w których ma to miejsce zwrócimy uwagę w przyszłości.

Rozkład entropii KL dla rozkładów punktowych:

Ze względu na $I_F = \sum_{n=1}^N I_{Fn}$, (2.186), z porównania (2.201) z (2.199) wynika dodatkowo, że entropia względna rozkładów $P(y|\theta)$ oraz $P(y|\theta + \Delta\theta)$ jest sumą entropii względnych odpowiadających im rozkładów punktowych $p_n(\mathbf{y}_n|\theta)$ oraz $p_n(\mathbf{y}_n|\theta + \Delta\theta)$:

$$S_H[P(y|\theta) | P(y|\theta + \Delta\theta)] = \sum_{n=1}^N S_H[p_n(\mathbf{y}_n|\theta) | p_n(\mathbf{y}_n|\theta + \Delta\theta)] \quad \text{dla dowolnego } \Delta\theta. \quad (2.202)$$

Ponieważ warunek $\Delta\theta \rightarrow 0$ nie był wykorzystywany w otrzymaniu (2.202) z porównania (2.199) oraz (2.201), zatem związek ten jest słuszny dla dowolnego $\Delta\theta$. Natomiast należy pamiętać, że rozkłady punktowe w definicji funkcji wiarygodności są *nieskorelowane* dla różnych n , dlatego też związek (2.202) jest słuszny tylko w tym przypadku.

2.7 Pojęcie kanału informacyjnego

Niech pierwotna zmienna losowa Y przyjmuje wartości wektorowe $\mathbf{y} \in \mathcal{Y}$. Wektor \mathbf{y} może być np. wektorem położenia. Zatem, aby wprowadzony opis był wystarczająco ogólny, wartości $\mathbf{y} \equiv (\mathbf{y}^\nu)$ mogą posiadać np. indeks wektorowy ν . Wartości te są realizowane zgodnie z łącznym rozkładem $p(\mathbf{y}|\Theta)$ właściwym dla badanego układu.

Rozważmy N -wymiarową próbę. Oznaczmy przez $\mathbf{y} \equiv (\mathbf{y}_n)_{n=1}^N = (\mathbf{y}_1, \dots, \mathbf{y}_N)$ dane będące realizacjami próby $\tilde{Y} = (Y_1, Y_2, \dots, Y_N)$ dla pierwotnej zmiennej Y , gdzie $\mathbf{y}_n \equiv (\mathbf{y}_n^\nu)$ oznacza n -tą wektorową obserwację w próbce ($n = 1, 2, \dots, N$). Rozkład łączny próby jest określony przez $P(\mathbf{y}|\Theta)$.

Dodatkowy indeks parametru: Podobnie, również parametry rozkładu mogą mieć dodatkowy indeks. Niech indeks α określa pewną dodatkową współrzędną wektorową parametru θ_i , gdzie jak w poprzednich rozdziałach $i = 1, 2, \dots, d$. Zatem wektor parametrów ma teraz postać:

$$\Theta = (\theta_1, \theta_2, \dots, \theta_d), \quad \text{gdzie} \quad \theta_i = (\theta_i^\alpha), \quad \alpha = 1, 2, \dots. \quad (2.203)$$

Wariancja estymatora $\hat{\theta}_i^\alpha$ parametru θ_i^α ma postać:

$$\sigma^2(\hat{\theta}_i^\alpha) = \int_{\mathcal{B}} d\mathbf{y} P(\mathbf{y}|\Theta) \left(\hat{\theta}_i^\alpha(\mathbf{y}) - \theta_i^\alpha \right)^2, \quad (2.204)$$

gdzie $\hat{\theta}_i^\alpha(\mathbf{y})$ jest estymatorem parametru θ_i^α , a całkowanie przebiega po całej przestrzeni próby \mathcal{B} , tzn. po wszystkich możliwych realizacjach \mathbf{y} .

Według Rozdziału 2.3.2.2, dla każdego wyróżnionego parametru θ_i^α wariancja $\sigma^2(\hat{\theta}_i^\alpha)$ jego estymatora (2.221) jest związana z informacją Fishera $I_{Fi\alpha}$ parametru θ_i^α poprzez nierówność (2.138):

$$\sigma^2(\hat{\theta}_i^\alpha) \geq I_F^{i\alpha} \geq \frac{1}{I_{Fi\alpha}}, \quad (2.205)$$

gdzie $I_F^{i\alpha} \equiv I_F^{i\alpha, i\alpha}$ jest dolnym ograniczeniem RC dla parametru θ_i^α w przypadku wieloparametrowym, natomiast:

$$I_{Fi\alpha} \equiv I_{Fi\alpha, i\alpha} \quad (2.206)$$

jest pojemnością informacyjną w *pojedynczym kanale informacyjnym* (i, α) czyli *informacją Fishera dla parametru θ_i^α* . Zgodnie z (2.183) jest ona równa²⁸:

$$I_{Fi\alpha} \equiv I_F(\theta_{i\alpha}) = \int_{\mathcal{B}} dy P(y|\Theta) \left(\frac{\partial \ln P(y|\Theta)}{\partial \theta_i^\alpha} \right)^2. \quad (2.207)$$

Informacja Stama: Wielkość $\frac{1}{\sigma^2(\hat{\theta}_i^\alpha)}$ odnosi się do pojedynczego kanału (i, α) . Sumując ją po indeksach α oraz i otrzymujemy tzw. informację Stama I_S [27]:

$$0 \leq I_S \equiv \sum_i \sum_\alpha \frac{1}{\sigma^2(\hat{\theta}_i^\alpha)}, \quad (2.208)$$

która jest skalarną miarą jakości jednoczesnej estymacji we wszystkich kanałach informacyjnych. *Informacja Stama jest z definicji zawsze wielkością nieujemną.*

Pojemność informacyjna I : W końcu, sumując lewą i prawą stronę (2.205) po indeksach α oraz i otrzymujemy następującą nierówność dla I_S :

$$0 \leq I_S \equiv \sum_i \sum_\alpha \frac{1}{\sigma^2(\hat{\theta}_i^\alpha)} \leq \sum_i \sum_\alpha I_{Fi\alpha} =: I \equiv C, \quad (2.209)$$

gdzie C , oznaczana dalej jako I , nazywana jest *pojemnością informacyjną układu*. Zgodnie z (2.209) i (2.207) jest ona równa:

$$I = \sum_i \sum_\alpha I_{Fi\alpha} = \sum_i \int_{\mathcal{B}} dy P(y|\Theta) \sum_\alpha \left(\frac{\partial \ln P(y|\Theta)}{\partial \theta_i^\alpha} \right)^2. \quad (2.210)$$

Jak się okaże, pojemność informacyjna I jest najważniejszym pojęciem statystyki leżącym u podstaw np. członów kinetycznych różnych modeli teorii pola. Jest ona uogólnieniem pojęcia informacji Fishera dla przypadku pojedynczego, skalarne parametru na przypadek wieloparametrowy.

2.7.1 Pojemność informacyjna dla zmiennej losowej położenia

Zawężmy obszar analizy do szczególnego przypadku, gdy interesującym nas oczekiwanym parametrem jest wartość oczekiwana zmiennej położenia układu Y :

$$\theta \equiv E(Y) = (\theta^\nu), \quad \text{gdzie} \quad \theta^\nu = \int_{\mathcal{Y}} d\mathbf{y} p(\mathbf{y}) \mathbf{y}^\nu. \quad (2.211)$$

Wtedy N -wymiarowa próbka $y \equiv (\mathbf{y}_n)_{n=1}^N = (\mathbf{y}_1, \dots, \mathbf{y}_N)$ jest realizacją próby \tilde{Y} dla *położeń układu*, a wartość oczekiwana θ_n położenia układu w n -tym punkcie (tzn. pomiarze) próby wynosi:

$$\theta_n \equiv E(Y_n) = (\theta_n^\nu), \quad \text{gdzie} \quad \theta_n^\nu = \int_{\mathcal{B}} dy P(y|\Theta) \mathbf{y}_n^\nu. \quad (2.212)$$

²⁸Porównaj przejście od (2.182) do (2.183).

Liczba oczekiwanych parametrów: Gdy, jak to ma miejsce w rozważanym przypadku, jedynym parametrem rozkładu, który nas interesuje jest wartość oczekiwana położenia $\theta_n \equiv (\theta_n^\nu)$, gdzie $n = 1, 2, \dots, N$ jest indeksem próby, wtedy parametr wektorowy $\Theta = (\theta_n)_{n=1}^N$. Zatem liczba parametrów θ_n pokryła się z wymiarem próby N , a indeksy parametru θ_i^α są następujące: $i \equiv n$, gdzie $n = 1, 2, \dots, N$, oraz $\alpha \equiv \nu$, gdzie ν jest indeksem wektorowym współrzędnej \mathbf{y}_n^ν . Oznacza to, że wymiar parametru Θ jest taki sam jak wymiar przestrzeni próby \mathcal{B} .

Współrzędne kowariantne i kontrawariantne: Rozważania obecnego Rozdziału jak i innych części skryptu są związane z analizą przeprowadzaną w czasoprzestrzeni Minowskiego. Dlatego koniecznym okazuje się rozróżnienie pomiędzy współrzędnymi kowariantnymi $\mathbf{y}_{n\nu}$ i kontrawariantnymi \mathbf{y}_n^μ . Związek pomiędzy nimi, tak dla wartości losowego wektora położenia jak i dla odpowiednich wartości oczekiwanych, jest następujący:

$$\mathbf{y}_{n\nu} = \sum_{\mu=0}^3 \eta_{\nu\mu} \mathbf{y}_n^\mu, \quad \theta_{n\nu} = \sum_{\mu=0}^3 \eta_{\nu\mu} \theta_n^\mu, \quad (2.213)$$

gdzie $(\eta_{\nu\mu})$ jest tensorem metrycznym przestrzeni \mathcal{Y} . W przypadku wektorowego indeksu Minkowskiego $\nu = 0, 1, 2, 3, \dots$ przyjmujemy następującą postać tensora metrycznego:

$$\eta_{\nu\mu} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & \ddots \end{pmatrix} \quad (2.214)$$

lub w skrócie $(\eta_{\nu\mu}) = \text{diag}(1, -1, -1, -1, \dots)$. Symbol “diag” oznacza macierz diagonalną z niezerowymi elementami na przekątnej głównej oraz zerami poza nią. Natomiast dla Euklidesowego indeksu wektorowego $\nu = 1, 2, 3, \dots$, tensor metryczny ma postać:

$$(\eta_{\nu\mu}) = \text{diag}(1, 1, 1, \dots) . \quad (2.215)$$

Założenie o niezależności p_n od θ_m dla $m \neq n$: W rozważaniach niniejszego skryptu zmienne Y_n próby \tilde{Y} są niezależne (tzn. zakładamy, że pomiary dla $m \neq n$ są w próbie niezależne). Oznacza to również, że wartość oczekiwana położenia $\theta_m = \int dy P(y|\Theta) \mathbf{y}_m$, nie ma wpływu na rozkład $p_n(\mathbf{y}_n|\theta_n)$ dla indeksu próby $m \neq n$. Wtedy dane są generowane zgodnie z punktowymi rozkładami spełniającymi warunek:

$$p_n(\mathbf{y}_n|\Theta) = p_n(\mathbf{y}_n|\theta_n), \quad \text{gdzie } n = 1, \dots, N, \quad (2.216)$$

a wiarygodność próby jest iloczynem:

$$P(y|\Theta) = \prod_{n=1}^N p_n(\mathbf{y}_n|\theta_n). \quad (2.217)$$

Pojemność informacyjna kanału: Dla parametru położenia (czaso)przestrzennego, pojemność kanału informacyjnego I dla układu, zdefiniowana ogólnie w (2.210), przyjmuje postać²⁹:

$$I = \sum_{n=1}^N I_{Fn} , \quad (2.219)$$

gdzie

$$\begin{aligned} I_{Fn} &\equiv I_{Fn}(\theta_n) = \int_{\mathcal{B}} dy P(y|\Theta) (\nabla_{\theta_n} \ln P(y|\Theta) \cdot \nabla_{\theta_n} \ln P(y|\Theta)) \\ &= \int_{\mathcal{B}} dy P(y|\Theta) \sum_{\nu, \mu=(0),1,2,\dots} \eta^{\nu\mu} \left(\frac{\partial \ln P(y|\Theta)}{\partial \theta_n^\nu} \frac{\partial \ln P(y|\Theta)}{\partial \theta_n^\mu} \right) . \end{aligned} \quad (2.220)$$

Tensor $(\eta^{\nu\mu}) = \text{diag}(1, -1, -1, -1)$ jest tensorem dualnym do $(\eta_{\nu\mu})$, tzn. $\sum_{\mu=0}^3 \eta_{\nu\mu} \eta^{\mu\gamma} = \delta_\nu^\gamma$, przy czym δ_ν^γ jest deltą Kroneckera, a $\nabla_{\theta_n} \equiv \frac{\partial}{\partial \theta_n} = \sum_\nu \frac{\partial}{\partial \theta_n^\nu} d\mathbf{y}_n^\nu$. W związku (2.220) “ \cdot ” oznacza iloczyn wewnętrzny zdefiniowany przez tensor metryczny $(\eta^{\nu\mu})$.

Uwaga o nefaktoryzowalności czasoprzestrzennych indeksów położenia: W pomiarze wybranej ν -tej współrzędnej położenia nie można wykluczyć odchyleń (fluktuacji) wartości współrzędnych do niej ortogonalnych. Oznacza to, że wartość oczekiwana ν -tej współrzędnej położenia nie jest w (2.211) liczona z jakiegoś rozkładu typu $p(\mathbf{y}^\nu)$, lecz musi być liczona z łącznego rozkładu $p(\mathbf{y})$ dla wszystkich współrzędnych \mathbf{y}^ν . W konsekwencji, w przypadku zmiennych położenia przestrzennego i ich parametrów naturalnych określonych w (2.211), całkowanie w (2.220) nie może zostać sfaktoryzowane ze względu na współrzędną wektorową ν .

Zgodnie z powyższą uwagą, wariancja estymatora $\hat{\theta}_n(y)$ parametru θ_n powinna przyjąć postać:

$$\begin{aligned} \sigma^2(\hat{\theta}_n) &= \int_{\mathcal{B}} dy P(y|\Theta) (\hat{\theta}_n(y) - \theta_n) \cdot (\hat{\theta}_n(y) - \theta_n) \\ &= \int_{\mathcal{B}} dy P(y|\Theta) \sum_{\nu, \mu=(0),1,2,\dots} \eta_{\nu\mu} (\hat{\theta}_n^\nu(y) - \theta_n^\nu) (\hat{\theta}_n^\mu(y) - \theta_n^\mu) . \end{aligned} \quad (2.221)$$

Rozważania przedstawione na końcu Rozdziału 2.3.2.2 oznaczają, że ze względu na (2.216), dla każdego wyróżnionego parametru θ_n wariancja $\sigma^2(\hat{\theta}_n)$ jego estymatora (2.221) w jego kanale informacyjnym jest związana z informacją Fishera $I_{Fn} = I_{Fn}(\theta_n)$ parametru θ_n poprzez nierówność (2.139):

$$\frac{1}{\sigma^2(\hat{\theta}_n)} \leq \frac{1}{I_{Fn}} \leq I_{Fn} \quad \text{gdzie } n = 1, 2, \dots, N , \quad (2.222)$$

będącą uogólnieniem nierówności informacyjnej Rao-Cramera (2.115), gdzie I_F^n jest DORC dla parametru θ_n .

²⁹Lokalne własności funkcji wiarygodności $P(y|\Theta)$ opisuje obserwowana macierz informacji Fishera:

$$\mathbf{iF} = \left(-\frac{\partial^2 \ln P(\Theta)}{\partial \theta_n^\nu \partial \theta_n^\mu} \right) , \quad (2.218)$$

która pozostaje symetryczna i dodatnio określona. Jak wiemy (por. Rozdział 2.2) jej wartość oczekiwana na \mathcal{B} zadaje geometryczną strukturę nazywaną metryką Rao-Fishera na przestrzeni statystycznej S [6].

Uwaga o zmiennych Fisher’owskich: Faktu *zależności statystycznej* zmiennych położenia przestrzennego dla różnych indeksów ν nie należy mylić z posiadaną przez nie *niezależnością analityczną*, która oznacza, że zmienne Y są tzw. zmiennymi Fisher’owskimi, dla których:

$$\frac{\partial \mathbf{y}^\nu}{\partial \mathbf{y}^\mu} = \delta_\mu^\nu . \quad (2.223)$$

Związek pomiędzy informacją Stama i pojemnością I : Wielkość $\frac{1}{\sigma^2(\hat{\theta}_n)}$ odnosi się do estymacji w n -tym kanale informacyjnym. Sumując po indeksie n otrzymujemy *informację Stama* I_S [27, 8]:

$$0 \leq I_S \equiv \sum_{n=1}^N \frac{1}{\sigma^2(\hat{\theta}_n)} =: \sum_{n=1}^N I_{Sn} . \quad (2.224)$$

Ponieważ $\theta_n = (\theta_n^\nu)$ jest parametrem wektorowym, zatem $I_{Sn} = \frac{1}{\sigma^2(\hat{\theta}_n)}$ jest informacją Stama czasoprzestrzennych kanałów dla n -tego pomiaru w próbie³⁰.

W końcu, sumując lewą i prawą stronę w (2.222) względem indeksu n i biorąc pod uwagę (2.208), zauważamy, że I_S spełnia nierówność:

$$0 \leq I_S \equiv \sum_{n=1}^N I_{Sn} \leq \sum_{n=1}^N I_{Fn} = I , \quad (2.225)$$

gdzie I jest pojemnością kanału informacyjnego (2.219). Nierówność (2.225) jest minimalnym uogólnieniem “jednokanałowej” nierówności Rao-Cramér’a (2.222), potrzebnym z punktu widzenia przeprowadzanego pomiaru.

Z punktu widzenia modelowania fizycznego, pojemność kanału informacyjnego I jest najważniejszym pojęciem statystycznym, leżącym u podstaw członów kinematycznych [8] różnych modeli teorii pola. Zgodnie z (2.208) okazało się, że zarówno dla metryki Euklidesowej (2.215) jak i metryki Minkowskiego (2.214), estymacja jest wykonywana dla dodatniej informacji I_S . Zatem z (2.209) wynika, że I jest również nieujemna. W Rozdziale 4.3.1 według (4.21) okaże się, że I jest nieujemnie zdefiniowana dla teorii pola dla cząstek, które mają nieujemny kwadrat masy [30]. Chociaż w każdym szczególnym modelu teorii pola z przestrzenią Minkowskiego fakt ten powinien zostać sprawdzony, to z punktu widzenia teorii estymacji jest jasne, że:

$$\sigma^2(\hat{\theta}_n) \geq 0 . \quad (2.226)$$

Sytuacja ta ma zawsze miejsce dla *procesów przyczynowych*.

³⁰Uwzględnienie metryki Minkowskiego w definicji informacji Stama można zrozumieć również jako konsekwencję ogólnego wskazania przy liczeniu średniej kwadratowej wielkości mierzalnej w dowolnej metryce Euklidesowej. W sytuacji gdy obok indeksów przestrzennych x_i , $i = 1, 2, 3$, występuje indeks czasowy t , należy w rachunkach w czterowymiarowej czasoprzestrzeni Euklidesowej uwzględnić we współrzędnych przestrzennych jednostkę urojoną i , *łącznie z uwzględnieniem tego faktu w prawie propagacji błędów*. W związku z tym w oryginalnej analizie EFI Friedena-Soffera, zmienna losowa czterowektora położenia oraz jej wartość oczekiwana mają odpowiednio postać $(Y_0 = cT, i\vec{Y})$ oraz $(\theta_0 = cE_\Theta(T), \vec{\theta} = iE(\vec{Y}))$, co nie zmienia rezultatów analizy zawartej w skrypcie, odnoszącej się do równań mechaniki falowej oraz termodynamiki. Nie zmienia to również rezultatów analizy relatywistycznej mechaniki kwantowej [29]. Jednakże opis wykorzystujący metrykę Minkowskiego wydaje się autorowi skryptu korzystniejszy z punktu widzenia zrozumienia konstrukcji niepodzielnego ekperymentalnie kanału informacyjnego (por. Uwaga o indeksie próby).

Uwaga o indeksie próby i kanale pomiarowym: Indeks próby n jest najmniejszym indeksem kanału informacyjnego, w którym dokonywany jest pomiar. Tzn. gdyby indeks próby można było dodatkowo “zaindeksować”, np. indeksem czasoprzestrzennym, wyznaczając podkanały, to i tak nie można by dokonać pomiaru tylko w jednym z tak wyznaczonych podkanałów (nie dokonując go równocześnie w pozostałych podkanałach posiadających indeks próby n). *Kanał niepodzielny z punktu widzenia eksperymentu nazwijmy kanałem pomiarowym.*

Uwaga o analizie we fragmencie kanału pomiarowego: W przypadku ograniczenia analizy do fragmentu kanału pomiarowego należy się upewnić, czy pozostała w analizie część informacji Stama ma wartość dodatnią. Np. w przypadku zaniedbania czasowo zaindeksowanej części czasoprzestrzennego kanału pomiarowego, otrzymana nierówność Stama dla składowych przestrzennych ma postać:

$$0 \leq I_S = \sum_{n=1}^N I_{Sn} \leq \sum_{n=1}^N \int d\vec{y} P(\vec{y}|\vec{\Theta}) \sum_{i=1}^3 \frac{\partial \ln P(\vec{y}|\vec{\Theta})}{\partial \theta_{ni}} \frac{\partial \ln P(\vec{y}|\vec{\Theta})}{\partial \theta_{ni}} =: \sum_{n=1}^N I_{Fn} = I, \quad (2.227)$$

gdzie znak wektora oznacza, że analiza zarówno w przestrzeni próby jak i przestrzeni parametrów została obciążona do części przestrzennej zmiennych losowych i parametrów.

Uwaga o symetrii: Z punktu widzenia pomiaru, błąd estymacji składowych wchodzących jednocześnie w wyznaczenie długości czterowektora w n -tym kanale, jest niezależny od układu współrzędnych w przestrzeni Minkowskiego. Dlatego wielkość $I_{Sn} = 1/\sigma^2(\hat{\theta}_n)$ określona poprzez (2.224) oraz (2.221) jest dla tensora metrycznego (2.214) niezmiennicza ze względu na transformację Lorentz’a (pchnięcia i obroty). W przypadku tensora metrycznego (2.215) jest ona niezmiennicza ze względu na transformację Galileusza.

Jeśli chodzi o pojemność informacyjną I , to w przypadku niezależności pomiarów w próbie, jest ona również niezmiennicza ze względu na transformację Lorentz’a w przestrzeni z metryką Minkowskiego (czy transformację Galileusza w przestrzeni Euklidesowej), o ile niezmiennicze jest każde I_n .³¹ Warunki niezmienniczości I_{Sn} oraz I schodzą się, gdy w nierówności Rao-Cramera (2.222) osiągnięta jest równość.

Nierówność Rao-Cramera okazuje się niezmiennicza ze względu na podstawowe transformacje [8, 28]. Istotnie, zgodnie z powyższymi rozważaniami, właściwym pomiarem niezależnym od przyjętego układu współrzędnych jest pomiar kwadratu długości $\sum_{\nu=0}^3 y_\nu y^\nu$, a nie pojedynczej współrzędnej y^ν . Więcej na temat niezmienniczości DORC ze względu na przesunięcie, odbicie przestrzenne, obroty i transformację afiniczną oraz transformacje unitarne można znaleźć w [28].

Kryterium minimalizacji I ze względu na N : Na koniec zauważmy, że sumowanie w (2.220) przebiega od $n = 1$ do $n = N$. Każdy n -ty wyraz w sumie wnosi analityczny wkład jako stopień swobody

³¹Pojemność informacyjna:

$$I_n = \int_{\mathcal{B}} dy P(y|\Theta) \sum_{\nu=0}^3 \mathfrak{F}_{nn\nu}^\nu \quad (2.228)$$

jest niezmiennicza ze względu na gładkie odwracalne odwzorowania $Y \rightarrow X$, gdzie X jest nową zmienną [21]. Jest ona również niezmiennicza ze względu na odbicia przestrzenne i czasowe.

dla I . O ile dodane stopnie swobody nie wpływają na już istniejące, to ponieważ każdy, cały wyraz w sumie po n jest nieujemny, to informacja I ma tendencje do wzrostu wraz ze wzrostem N . Kryterium minimalizacji I ze względu na N posłużyło Friedenowi i Sofferowi jako dodatkowy warunek przy konstrukcji np. równań ruchu. Nie znaczy to, że modele z większym N zostały automatycznie wykluczone, tylko że im większe jest N tym więcej stopni swobody wchodzi do opisu obserwowanego zjawiska i opisywane zjawisko jest bardziej złożone. Zagadnienie to omówimy w przykładach, w dalszej części skryptu.

2.8 Pomiar układu w podejściu Friedena-Soffera

Jak w Rozdziale 2.7.1, rozważmy zmienną losową Y położenia układu, przyjmującą wartość y , która jest punktem zbioru \mathcal{Y} . Może to być punkt czasoprzestrzenny przestrzeni Minkowskiego, co ma miejsce w rozważaniach związanych z opisem układu np. w mechanice falowej. Wartości $y \equiv (y^\nu)_{\nu=0}^3 \in \mathcal{Y} \equiv \mathbb{R}^4$ są realizowane zgodnie z rozkładem $p(y)$ właściwym dla układu³².

Podstawowe założenie fizyczne podejścia Friedena-Soffera: Niech dane $y \equiv (y_n)_{n=1}^N = (y_1, \dots, y_N)$ są realizacjami próby dla położenia układu, gdzie $y_n \equiv (y_n^\nu)_{\nu=0}^3$. Zgodnie z założeniem zaproponowanym przez Frieden'a i Soffer'a [8], ich zebranie następuje przez sam układ w zgodzie z rozkładami gęstości prawdopodobieństwa, $p_n(y_n|\theta_n)$, gdzie $n = 1, \dots, N$.

Treść powyższego założenia fizycznego można wypowiedzieć następująco: *Układ próbkuje dostępną mu czasoprzestrzeń, "zbierając dane i dokonując analizy statystycznej", zgodnie z zasadami informacyjnymi (wprowadzonymi w Rozdziale 3).*

Przestrzenie statystyczne próby \mathcal{S} , punktowa \mathcal{S}_4 oraz $\mathcal{S}_{N \times 4}$: Niech rozważana przestrzeń jest czasoprzestrzenią Minkowskiego. Wtedy każdy z rozkładów $p_n(y_n|\theta_n)$ jest punktem modelu statystycznego $\mathcal{S}_4 = \{p_n(y_n|\theta_n)\}$ parametryzowanego przez naturalny parametr, tzn. przez wartość oczekiwaną $\theta_n \equiv (\theta_n^\nu)_{\nu=0}^3 = E(Y_n)$, jak w (2.212). Zbiór wartości $d = 4 \times N$ - wymiarowego parametru $\Theta = (\theta_n)_{n=1}^N$ tworzy współrzędne dla łącznego rozkładu $P(y|\Theta)$, będącego punktem na $d = 4 \times N$ -wymiarowej rozmiarowości, która jest (pod)przestrzenią statystyczną $\mathcal{S} \subset \Sigma(\mathcal{B})$ [6] (por. (2.34)):

$$\mathcal{S} = \{P_\Theta \equiv P(y|\Theta), \Theta \equiv (\theta_n)_{n=1}^N \in V_\Theta \subset \mathbb{R}^d\}, \quad (2.229)$$

określoną na przestrzeni próby \mathcal{B} . Jak wiemy z Rozdziału 2.2, rozmiarowość \mathcal{S} jest rodziną rozkładów prawdopodobieństwa parametryzowaną przez rzeczywistą, nie losową zmienną wektorową $\Theta \equiv (\theta_n)_{n=1}^N \in V_\Theta$, która w rozważanym przypadku parametru położenia, tworzy $N \times 4$ -wymiarowy lokalny układ współrzędnych. Zatem, ponieważ próba \tilde{Y} jest $N \times 4$ -wymiarową zmienną losową, więc wymiary przestrzeni próby \mathcal{B} i wektorowego parametru $\Theta \equiv (\theta_n)_{n=1}^N$ są takie same³³.

W przyszłości okaże się, że z powodu związku (2.216) analiza na przestrzeni statystycznej określonej przez (2.229) z parametrami $(\theta_n)_{n=1}^N$ tworzącymi $N \times 4$ -wymiarowy lokalny układ współrzędnych, efektywnie redukuje się do analizy na $\mathcal{S}_{N \times 4} \equiv \{\bigoplus_{n=1}^N p_n(y|\theta_n)\}$. Jednakże, ponieważ wartości parametru θ_n mogą się zmieniać od jednego punktu n próby do innego punktu n' , zatem nie może być

³²Przy wyprowadzaniu równań generujących rozkład w fizyce statystycznej y może być np. wartością energii ϵ układu [8] i wtedy $y \equiv \epsilon \in \mathcal{Y} \equiv \mathbb{R}$.

³³Jednakże przypomnijmy, że w ogólnym przypadku estymacji, wymiar wektora parametrów $\Theta \equiv (\theta_i)_{i=1}^d$ oraz wektora próby $y \equiv (y_n)_{n=1}^N$ może być inny.

ona w ogólności sprowadzona do analizy na \mathcal{S} poprzez samo przeskalowanie metryki Riemanna oraz (dualnej) koneksji na \mathcal{S}_4 przez czynnik N , jak to ma miejsce we wnioskowaniu pojawiającym się w statystyce klasycznej [6]. W przyszłości liczbę N parametrów θ_n , będącą wymiarem próby $y \equiv (\mathbf{y}_n)_{n=1}^N$, będziemy nazywali rangą pola.

Uwaga o podmiocie: Jednak pamiętajmy, że estymacji dokonuje tylko człowiek, zatem na metodę EFI należy patrzeć tylko jak na *pewien model analizy statystycznej*.

Interesująca nas statystyczna procedura estymacyjna dotyczy wnioskowania o $p_n(\mathbf{y}_n|\theta_n)$ na podstawie danych y z wykorzystaniem funkcji wiarygodności $P(y|\Theta)$. Załóżmy, że dane zbierane przez układ $y = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)$ są uzyskane niezależnie tak, że łączny rozkład prawdopodobieństwa dla próby faktoryzuje się na rozkłady brzegowe:

$$P(\Theta) \equiv P(y|\Theta) = \prod_{n=1}^N p_n(\mathbf{y}_n|\Theta) = \prod_{n=1}^N p_n(y_n|\theta_n), \quad (2.230)$$

gdzie w ostatniej równości skorzystano z założenia, że parametr θ_m dla $m \neq n$ nie ma wpływu na rozkład zmiennej Y_n .

Wstępne określenie postaci kinematycznej I : Centralna część pracy Frieden'a i Soffer'a jest związana z przejściem od pojemności informacyjnej I zadanej równaniem (2.220) oraz (2.219):

$$I = \sum_{n=1}^N I_n = \sum_{n=1}^N \int_{\mathcal{B}} dy P(y|\Theta) \sum_{\nu=0}^3 \left(\frac{\partial \ln P(y|\Theta)}{\partial \theta_{n\nu}} \frac{\partial \ln P(y|\Theta)}{\partial \theta_n^\nu} \right), \quad (2.231)$$

gdzie $dy := d^4\mathbf{y}_1 \dots d^4\mathbf{y}_N$ oraz $d^4\mathbf{y}_n = d\mathbf{y}_n^0 d\mathbf{y}_n^1 d\mathbf{y}_n^2 d\mathbf{y}_n^3$, do tzw. postaci kinematycznej wykorzystywanej w teorii pola oraz fizyce statystycznej.

Rachunek analogiczny jaki doprowadził z (2.183) do (2.186) wygląda teraz w skrócie następująco. Przekształćmy pochodną $\ln P$ w (2.231) do postaci:

$$\frac{\partial \ln P(y|\Theta)}{\partial \theta_{n\nu}} = \frac{\partial}{\partial \theta_{n\nu}} \sum_{n=1}^N \ln p_n(\mathbf{y}_n|\theta_n) = \sum_{n=1}^N \frac{1}{p_n(\mathbf{y}_n|\theta_n)} \frac{\partial p_n(\mathbf{y}_n|\theta_n)}{\partial \theta_{n\nu}}. \quad (2.232)$$

Pamiętając o unormowaniu każdego z rozkładów brzegowych, $\int_{\mathcal{Y}} d^4\mathbf{y}_n p_n(\mathbf{y}_n|\theta_n) = 1$, otrzymujemy postać pojemności informacyjnej:

$$I = \sum_{n=1}^N \int_{\mathcal{Y}} d^4\mathbf{y}_n \frac{1}{p_n(\mathbf{y}_n|\theta_n)} \sum_{\nu=0}^3 \left(\frac{\partial p_n(\mathbf{y}_n|\theta_n)}{\partial \theta_{n\nu}} \frac{\partial p_n(\mathbf{y}_n|\theta_n)}{\partial \theta_n^\nu} \right), \quad (2.233)$$

będącą uogólnieniem (2.186).

W końcu przejdźmy do amplitud $q_n(\mathbf{y}_n|\theta_n)$ określonych jak w (2.168):

$$p_n(\mathbf{y}_n|\theta_n) = q_n^2(\mathbf{y}_n|\theta_n). \quad (2.234)$$

Proste rachunki dają:

$$I = 4 \sum_{n=1}^N \int_{\mathcal{Y}} d^4\mathbf{y}_n \sum_{\nu=0}^3 \left(\frac{\partial q_n(\mathbf{y}_n|\theta_n)}{\partial \theta_{n\nu}} \frac{\partial q_n(\mathbf{y}_n|\theta_n)}{\partial \theta_n^\nu} \right), \quad (2.235)$$

czyli prawie kluczową postać pojemności informacyjnej dla rachunku metody EFI Friedena-Soffera. Jedyne co trzeba jeszcze zrobić, to przejść od przestrzeni statystycznej \mathcal{S} z bazą $(\theta_n)_{n=1}^N$ dla reprezentacji amplitud danych pomiarowych $\mathbf{y}_n \in \mathcal{Y}$, do przestrzeni amplitud przesunąć $\mathbf{x}_n := \mathbf{y}_n - \theta_n$ określonych na przestrzeni bazowej \mathcal{X} . Poświęćmy temu zagadnieniu Rozdział 3.3.

2.8.1 Przykład: Estymacja w fizycznych modelach eksponenialnych

W Rozdziale 2.2.2 wprowadzone zostało pojęcie dualnych affinicznych układów współrzędnych na przestrzeni statystycznej \mathcal{S} . Obecny rozdział poświęcony jest zastosowaniu modeli eksponenialnych (2.61) i szczególnej roli parametrów dualnych związanych z koneksją $\nabla^{(-1)}$. Rodzina modeli eksponenialnych wykorzystywana jest w teorii estymacji szeregu zagadnień fizycznych. Jej najbardziej znaną realizacją jest estymacja metodą maksymalnej entropii, sformułowana w poniższym twierdzeniu dla wymiaru próby $N = 1$.

Niech:

$$S_H(p) = - \int_{\mathcal{Y}} d\mathbf{y} p(\mathbf{y}|\Xi) \ln p(\mathbf{y}|\Xi) , \quad \forall p_{\Xi} \in \mathcal{S} \quad (2.236)$$

jest entropią Shannona stanu układu zadanego rozkładem $p(\mathbf{y}|\Xi)$ z parametrami Ξ , a $F_i(Y)$, $i = 1, 2, \dots, d$, układem niezależnych zmiennych losowych o określonych wartościach oczekiwanych:

$$\theta_i = \theta_i(\Xi) = E_{\Xi} [F_i(Y)] = \int_{\mathcal{Y}} d\mathbf{y} p(\mathbf{y}|\Xi) F_i(\mathbf{y}) , \quad i = 1, 2, \dots, d \quad , \quad \forall p_{\Xi} \in \mathcal{S} . \quad (2.237)$$

Z (2.237) widać, że $F_i(Y)$ są nieobciążonymi estymatorami parametrów θ_i , tzn.:

$$\hat{\theta}_i = F_i(Y) , \quad i = 1, 2, \dots, d . \quad (2.238)$$

Twierdzenie o stanie z maksymalną entropią (TME). Istnieje jednoznacznie określony unormowany stan układu posiadający maksymalną entropię, zadany następująco:

$$p_{\Xi} \equiv p(\mathbf{y}|\Xi) = Z^{-1} \exp \left(\sum_{i=1}^d \xi^i F_i(\mathbf{y}) \right) , \quad \forall p_{\Xi} \in \mathcal{S} \quad (2.239)$$

w bazie kanonicznej $\Xi \equiv (\xi^i)_{i=1}^d$ modelu eksponenialnego (2.61), gdzie stała normalizacyjna Z jest tzw. funkcją partycji (podziału) [31]:

$$Z \equiv Z(\Xi) = \int_{\mathcal{Y}} d\mathbf{y} \exp \left(\sum_{i=1}^d \xi^i F_i(\mathbf{y}) \right) . \quad (2.240)$$

Dowód: Zgodnie z twierdzeniem Lagrange'a wiemy, że maksimum warunkowe dla funkcji $S_H(p)$, przy dodatkowym warunku normalizacyjnym:

$$\int_{\mathcal{Y}} d\mathbf{y} p(\mathbf{y}|\Xi) = 1 \quad (2.241)$$

oraz d warunkach związanych z zadaniem wartości oczekiwanych (2.237), jest równoważne wyznaczeniu bezwarunkowego ekstremum funkcji:

$$\begin{aligned} S_{war}(p) &= \int_{\mathcal{Y}} d\mathbf{y} s_{war}(p) := S_H(p) - \xi^0 \left(1 - \int_{\mathcal{Y}} d\mathbf{y} p(\mathbf{y}|\Xi) \right) - \sum_{i=1}^d \xi^i \left(\theta_i - \int_{\mathcal{Y}} d\mathbf{y} p(\mathbf{y}|\Xi) F_i(\mathbf{y}) \right) \\ &= \int_{\mathcal{Y}} d\mathbf{y} p(\mathbf{y}|\Xi) \left(-\ln p(\mathbf{y}|\Xi) + \xi^0 + \sum_{i=1}^d \xi^i F_i(\mathbf{y}) \right) - \xi^0 - \sum_{i=1}^d \xi^i \theta_i , \end{aligned} \quad (2.242)$$

ze względu na wariację $p(\mathbf{y}|\Xi)$, gdzie ξ^0 oraz $(\xi^i)_{i=1}^d$ są czynnikami Lagrange'a. Ponieważ $S_H(p)$ jest ściśle wklęsła [20], zatem otrzymujemy *maksimum, które jest wyznaczone jednoznacznie*.

Warunek ekstremizacji funkcjonału³⁴ $S_{war}(p)$ ze względu na $p(\mathbf{y}|\Xi)$, tzn. $\delta_{(p)}S_{war} = 0$, prowadzi do równania Eulera-Lagrange'a:

$$\frac{\partial}{\partial \mathbf{y}} \left(\frac{\partial S_{war}}{\partial \left(\frac{\partial p(\mathbf{y}|\Xi)}{\partial \mathbf{y}} \right)} \right) = \frac{\partial S_{war}}{\partial p(\mathbf{y}|\Xi)} , \quad (2.243)$$

gdzie zgodnie z (2.242) postać funkcji podcałkowej s_{war} wynosi:

$$s_{war} = p(\mathbf{y}|\Xi) \left(-\ln p(\mathbf{y}|\Xi) + \xi^0 + \sum_{i=1}^d \xi^i F_i(\mathbf{y}) \right) . \quad (2.244)$$

Podstawiając s_{war} do (2.243) otrzymujemy:

$$-\ln p(\mathbf{y}|\Xi) + \xi^0 + \sum_{i=1}^d \xi^i F_i(\mathbf{y}) - 1 = 0 . \quad (2.245)$$

Równanie (2.245) daje szukaną postać rozkładu $p(\mathbf{y}|\Xi)$ maksymalizującego entropię $S_H(p)$:

$$p(\mathbf{y}|\Xi) = A \exp \left(\sum_{i=1}^d \xi^i F_i(\mathbf{y}) \right) , \quad \text{gdzie} \quad A = 1/\exp(1 - \xi^0) = \text{const.} , \quad (2.246)$$

który jak to widać z (2.61) jest typu eksponentialnego z parametrami kanonicznymi ξ^i , $i = 1, 2, \dots, d$ oraz $C(\mathbf{y}) = 0$. Ponadto z warunku normalizacji (2.241) otrzymujemy $A = Z^{-1}$, gdzie Z jest funkcją partycji (2.240). c.n.d.

Parametry dualne. Po rozpoznaniu, że model maksymalizujący entropię stanu układu jest modelem eksponentialnym (2.239) w parametryzacji kanonicznej Ξ , możemy (2.239) zapisać w postaci (2.61):

$$p_{\Xi} \equiv p(\mathbf{y}|\Xi) = \exp \left[\sum_{i=1}^d \xi^i F_i(\mathbf{y}) - \psi(\Xi) \right] , \quad \forall p_{\Xi} \in \mathcal{S} , \quad (2.247)$$

gdzie

$$Z(\Xi) = \exp [\psi(\Xi)] . \quad (2.248)$$

Ponieważ zgodnie z (2.63), $\psi(\Xi) = \ln \int_{\mathcal{Y}} d\mathbf{y} \exp \left[\sum_{i=1}^d \xi^i F_i(\mathbf{y}) \right]$, zatem z (2.237) oraz wykorzystując (2.247), otrzymujemy:

$$\theta_i = \partial_{\xi^i} \psi(\Xi) , \quad i = 1, 2, \dots, d , \quad \forall p_{\Xi} \in \mathcal{S} , \quad (2.249)$$

gdzie skorzystano z oznaczenia $\partial_{\xi^i} \equiv \partial/\partial \xi^i$.

Korzystając z $\partial_{\xi^j} \partial_{\xi^i} \ln p(\mathbf{y}|\Xi) = -\partial_{\xi^j} \partial_{\xi^i} \psi(\Xi)$, (2.66), oraz $g_{ij}^{\xi} = -E_{\Xi}(\partial_{\xi^i} \partial_{\xi^j} \ln p(\mathbf{y}|\Xi))$, (2.37), otrzymujemy również:

$$g_{ij}^{\xi} = \partial_{\xi^i} \partial_{\xi^j} \psi(\Xi) , \quad i, j = 1, 2, \dots, d , \quad i, j = 1, 2, \dots, d , \quad \forall p \in \mathcal{S} . \quad (2.250)$$

³⁴Funkcjonał, w tym przypadku $S_{war}(p)$, jest liczbą, której wartość zależy od funkcji $p(\mathbf{y}|\Xi)$.

Dualne układy modelu maksymalizującego entropię: Z (2.64) wiemy, że układ współrzędnych $(\xi^i)_{i=1}^d$ jest $\alpha = 1$ - afinicznym układem współrzędnych modelu eksponentialnego. Zatem z (2.249) oraz (2.250) wynika, że $(\theta_i)_{i=1}^d$ jest $\alpha = (-1)$ - afinicznym układem współrzędnych modelu eksponentialnego dualnym do $(\xi^i)_{i=1}^d$.

Parametry θ_i , $i = 1, 2, \dots, d$, nazywamy z przyczyn podanych powyżej *parametrami dualnymi* do $(\xi^i)_{i=1}^d$ lub *parametrami oczekiwanymi* modelu statystycznego \mathcal{S} .

Estymacja na DORC: Korzystając z definicji (2.36) metryki Rao-Fishera, z postaci (2.247) rozkładu eksponentialnego oraz z (2.249), otrzymujemy postać macierzy informacyjnej I_F w parametryzacji kanonicznej Ξ :

$$\begin{aligned} I_{Fij}(\Xi) \equiv g_{ij}^\Xi &= E_\Xi [\partial_{\xi^i} \ln p(Y|\Xi) \partial_{\xi^j} \ln p(Y|\Xi)] \\ &= E_\Xi [(F_i(Y) - \theta_i)(F_j(Y) - \theta_j)] , \quad i, j = 1, 2, \dots, d , \quad \forall p_\Xi \in \mathcal{S} . \end{aligned} \quad (2.251)$$

Niech $\hat{\Theta} \equiv (\hat{\theta}_i)_{i=1}^d$ są estymatorami parametrów $\Theta \equiv (\theta_i)_{i=1}^d$ bazy dualnej do Ξ . Ponieważ z (2.238) mamy $\hat{\theta}_i = F_i(Y)$, $i = 1, 2, \dots, d$, zatem po prawej stronie (2.251) stoją elementy macierzy kowariancji $V_\Xi(\hat{\Theta})$ estymatorów $\hat{\Theta}$:

$$V_{\Xi ij}(\hat{\Theta}) = E_\Xi [(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)] , \quad i, j = 1, 2, \dots, d , \quad \forall p_\Xi \in \mathcal{S} . \quad (2.252)$$

Równość (2.251) można więc zapisać następująco:

$$V_\Xi(\hat{\Theta}) = I_F(\Xi) , \quad \forall p_\Xi \in \mathcal{S} . \quad (2.253)$$

Ponieważ macierze informacyjne w bazach dualnych są względem siebie odwrotne, tzn. $I_F(\Xi) = I_F^{-1}(\Theta)$, (2.97), oraz macierz kowariancji określonych zmiennych losowych (w tym przypadku $\hat{\Theta}$) nie zależy od bazy w \mathcal{S} :

$$V_\Theta(\hat{\Theta}) = V_\Xi(\hat{\Theta}) , \quad \forall p \in \mathcal{S} , \quad (2.254)$$

więc z (2.253) otrzymujemy następujący związek:

$$V_\Theta(\hat{\Theta}) = I_F^{-1}(\Theta) , \quad \forall p_\Theta \in \mathcal{S} . \quad (2.255)$$

Wniosek: Ze względu na Twierdzenie Rao-Cramera (2.142) powyższy warunek oznacza, że estymacja parametrów dualnych jest dla modeli eksponentialnych, spełniających warunek maksymalnej entropii dokonywana na DORC.

Przykład dualnego układu współrzędnych. Rozkład normalny: Z Rozdziału 2.2, wzór (2.69), wiemy, że rozkład normalny jest typem modelu eksponentialnego z $C(\mathbf{y}) = 0$. Oznacza to, że może się on pojawić jako rezultat estymacji spełniającej założenia TME. Korzystając z postaci rozkładu normalnego oraz z (2.237) i (2.69) otrzymujemy dualne parametry tego modelu:

$$\begin{aligned} \theta_1 &\equiv \theta_1(\xi^1, \xi^2) = \int_{\mathcal{Y}} d\mathbf{y} p(\mathbf{y}|\Xi) \mathbf{y} = -\frac{\xi^1}{2\xi^2} = \mu , \\ \theta_2 &\equiv \theta_2(\xi^1, \xi^2) = \int_{\mathcal{Y}} d\mathbf{y} p(\mathbf{y}|\Xi) \mathbf{y}^2 = \frac{(\xi^1)^2 - 2\xi^2}{4(\xi^2)^2} = \mu^2 + \sigma^2 . \end{aligned} \quad (2.256)$$

Ponadto estymatory $\hat{\theta}_1 = F_1(Y) = Y$ oraz $\hat{\theta}_2 = F_2(Y) = Y^2$ są niezależne. Sprawdźmy, że zachodzi warunek konieczny ich niezależności, a mianowicie brak korelacji:

$$\begin{aligned} E_{\Xi} \left[(\hat{\theta}_1 - \theta_1) (\hat{\theta}_2 - \theta_2) \right] &= \int_{\mathcal{Y}} d\mathbf{y} p(\mathbf{y}|\Xi) (F_1(\mathbf{y}) - \theta_1) (F_2(\mathbf{y}) - \theta_2) \\ &= \frac{1}{\sqrt{2\pi}\sigma^2} \int_{\mathcal{Y}} d\mathbf{y} e^{-\frac{(\mathbf{y}-\mu)^2}{2\sigma^2}} (\mathbf{y} - \mu) [\mathbf{y}^2 - (\mu^2 + \sigma^2)] \\ &= \frac{1}{\sqrt{2\pi}\sigma^2} \left(\int_{\mathcal{Y}} d\mathbf{y} e^{-\frac{(\mathbf{y}-\mu)^2}{2\sigma^2}} (\mathbf{y} - \mu) [\mathbf{y}^2 - \mu^2] - \sigma^2 \int_{\mathcal{Y}} d\mathbf{y} e^{-\frac{(\mathbf{y}-\mu)^2}{2\sigma^2}} (\mathbf{y} - \mu) \right) = 0, \end{aligned} \quad (2.257)$$

gdzie w drugiej linii skorzystano z postaci (1.16) rozkładu normalnego, a w ostatniej z zerowania się obu całek z osobna.

Wniosek: Z powyższych ogólnych rozważań wnioskujemy więc, że estymacja z wieloparametrowym rozkładem normalnym spełnia DORC. Poprzednio w Rozdziale 2.1, w wyniku bezpośredniego rachunku dla jednoparametrowej estymacji wartości oczekiwanej μ , otrzymaliśmy w (2.18) ten sam wynik.

Jednakże dla rozkładu normalnego, w którym chcielibyśmy dokonać jednoczesnej estymacji parametrów μ oraz σ^2 , pojawiłby się problem z zastosowaniem TRC wynikający z faktu, że dla skończonego wymiaru próby N estymator $\hat{\sigma}^2$ jest obciążony. Obecnie wiemy, że parametrami oczekiwanymi, którymi należy się posłużyć aby zastosować TR i przekonać się, że model normalny spełnia DORC są μ oraz suma $\mu^2 + \sigma^2$ (zamiast σ^2). Chociaż powyższy rachunek został przeprowadzony dla $N = 1$, jednak wniosek dla dowolnego N nie ulega zmianie.

Przykład dualnego układu współrzędnych. Rozkład standardowy eksponentialny: TME ma swoją reprezentację w fizyce statystycznej. Otóż stan w równowadze termicznej, który maksymalizuje termodynamiczną entropię Boltzmanna $S_B(p) := k_B S_H(p)$, gdzie k_B jest dodatnią stałą Boltzmanna, posiada przy warunku $E_{\xi}[E] = \bar{\epsilon}$ nałożonym na wartość oczekiwaną (zmiennej losowej) energii E cząstki gazu, rozkład Boltzmanna:

$$p(\epsilon|\xi) = \frac{1}{Z} e^{-\frac{2\epsilon}{3k_B T}}, \quad (2.258)$$

gdzie ϵ jest realizacją E , a T jest (entropijną) temperaturą (por. Dodatek 7.4). Rozkład (2.258) jest standardowym rozkładem eksponentialnym z $C(\epsilon) = 0$, (2.71), dla wymiaru próby $N = 1$. Zgodnie z oznaczeniami wprowadzonymi w (2.71), jeden parametr kanoniczny ξ , jedna funkcja $\hat{\theta}_{\epsilon} = F(E)$ będąca estymatorem parametru oczekiwanego θ_{ϵ} oraz potencjał $\psi(\xi)$ układu współrzędnych, mają postać:

$$F(E) = E, \quad \xi = -\frac{2}{3k_B T}, \quad \theta_{\epsilon} = \bar{\epsilon}, \quad \psi(\xi) = \ln\left(-\frac{1}{\xi}\right) = \ln Z. \quad (2.259)$$

Do rozkładu Boltzmanna powrócimy w Rozdziale 5.1.3, gdzie wyprowadzimy go odwołując się do wspomnianych w Rozdziale 2.2.3 zasad informacyjnych.

Kolejne przykłady zastosowania estymacji eksponentialnej z bazą dualną Θ : Estymacja tego typu znajduje swoje zastosowanie wtedy, gdy mikrostan układu jest po krótkim okresie czasu zastąpiony makrostanem, co oznacza estymację stanu układu poprzez pewien oszacowujący stan otrzymany metodą maksymalnej entropii na rozmaitości modelu eksponentialnego (np. dla wolno zmieniających się) zmiennych makroskopowych [21]. Przykładami realizacji tej procedury estymacyjnej są:

- **Metoda analizy nieliniowej dynamiki** Kossakowskiego i Ingardena, którzy zrealizowali powyższą procedurę, dokonując ciągłego rzutowania mikrostanu układu na łatwiejsze w opisie makrostany układu, leżące na rozmaitości stanów eksponentialnych. Z zaproponowanej analizy statystycznej wynika możliwość realizacji nieliniowej dynamiki układu, opisaną jako konsekwencja optymalnej *estymacji stanu układu*, pojawiającego się po upływie każdego kolejnego odstępu czasu (w którym dynamika układu przebiegała w sposób liniowy), *stanem leżącym na rozmaitości eksponentialnej* [21].

- **Model Onsagera** realizujący tego typu estymację w badaniu zjawiska przepływów energii lub masy, w sytuacji, gdy są one liniowymi funkcjami bodźców (pełniących rolę parametrów) wywołujących taki przepływ. Teoria Onsagera ma zastosowanie do zjawisk mających charakter *procesów quasistatystycznych*. Zatem stosuje się ona do sytuacji, gdy materiał, w którym zachodzi zjawisko jest w lokalnej równowadze, tzn. związki zachodzące lokalnie i w tej samej chwili czasu pomiędzy własnościami cieplnymi i mechanicznymi materiału są takie same, jak dla jednorodnego układu znajdującego się w równowadze termodynamicznej. W ramach jego teorii sformułowano zasady wariacyjne dla opisu liniowej termodynamiki procesów nieodwracalnych.

Podsumowanie: Z powyższej analizy wynika, że metoda maksymalnej entropii dla nieobciążonych estymatorów $\hat{\theta}_i = F_i(Y)$, $i = 1, 2, \dots, d$, wykorzystuje $\alpha = (-1)$ – płaską bazę Θ dualną do $\alpha = (+1)$ – płaskiej bazy kanonicznej Ξ rozkładu eksponentialnego z $C(\mathbf{y}) = 0$. Ze względu na płaskość modelu eksponentialnego w bazie kanonicznej Ξ , (2.64), macierz informacyjna I_F w bazie dualnej Θ jest odwrotna do I_F w bazie Ξ , skąd w (2.255) przekonaliśmy się, że estymacja w bazie koneksji affinicznej $\nabla^{(-1)}$ przebiega na DORC. Oznacza to, że również dla wektorowego parametru oczekiwanego Θ jego estymacja jest *efektywna* w klasie entropijnych modeli eksponentialnych.

Jednak idąc dalej, dokładniejsza niż to wynika z TRC, *modelowa estymacja*, tzn. związana z konstrukcją nieobciążonych estymatorów parametrów, nie jest możliwa.

O tym co w kolejnej części skryptu: W kolejnej części skryptu zajmiemy się estymacją związaną z zasadami informacyjnymi nałożonymi na tzw. *fizyczną informację układu*.

Rozdział 3

Zasady informacyjne

3.1 Estymacja w statystyce klasycznej a estymacja fizyczna. Postawienie problemu

W dotychczasowej analizie przedstawiona została MNW w statystyce. Polega ona na estymacji parametrów pewnego zadanego rozkładu. Na przykład w analizie regresji na podstawie pewnej wcześniejszej wiedzy na temat zachowania się zmiennej objaśnianej, zakresu wartości jakie może przyjmować oraz jej charakteru (ciągła czy dyskretna) postulujemy warunkowy rozkład i model regresji, a następnie konstruujemy funkcję wiarygodności, którą maksymalizując otrzymujemy estymatory parametrów strukturalnych modelu. Opracowanie skutecznego algorytmu znajdowania estymatorów MNW oraz ich odchyłeń standardowych jest centralnym problemem np. w rutynowych aplikacjach służących do analizy uogólnionych regresyjnych modeli liniowych. W analizie tej najważniejszym wykorzystywanym algorytmem jest ogólny algorytm metody iteracyjnie ważonych najmniejszych kwadratów, a jedną z jego głównych analitycznych procedur jest procedura Newton-Raphson'a [13, 15]. Niech parametr wektorowy $\Theta \equiv (\theta_n)_{n=1}^N$ jest zbiorem wartości oczekiwanych zmiennej losowej położenia układu w N pomiarach, jak to przyjęliśmy w Rozdziale 2.7. Przypomnijmy więc, że MNW jest wtedy skoncentrowana na układzie N równań wiarygodności (1.8):

$$S(\Theta) |_{\Theta=\hat{\Theta}} \equiv \frac{\partial}{\partial \Theta} \ln P(\Theta) |_{\Theta=\hat{\Theta}} = 0 ,$$

których rozwiązanie daje N elementowy zbiór $\hat{\Theta} \equiv (\hat{\theta}_n)_{n=1}^N$ estymatorów parametrów. Tzn. układ równań wiarygodności tworzy N warunków na estymatory parametrów, które maksymalizują wiarygodność próbki.

Estymacja w fizyce musi się rozpocząć na wcześniejszym etapie. Wychodząc od zasad informacyjnych, którym poświęcony będzie kolejny rozdział, estymujemy odpowiednie dla opisywanego zagadnienia fizycznego równania ruchu, których rozwiązanie daje odpowiedni rozkład wraz z parametrami. Tak więc zastosowanie zasad informacyjnych nałożonych na funkcję wiarygodności zamiast MNW stanowi o podstawowej różnicy pomiędzy analizą statystyczną wykorzystywaną w konstrukcji modeli fizycznych, a statystyką klasyczną. Oczywiście oznacza to, że informacja Fishera zdefiniowana poprzednio na przestrzeni statystycznej \mathcal{S} musi zostać związana z bazową przestrzenią \mathcal{Y} przestrzeni próby, tak aby można ją wykorzystać do konstrukcji równań ruchu.

3.1.1 Strukturalna zasada informacyjna. Metoda EFI

Poniższe rozważania prezentują analizę, leżącą u podstaw strukturalnej zasady informacyjnej [10]. Ta zaś leży u podstaw metody estymacji statystycznej EFI zaproponowanej przez Friedena i Soffera [8]. Niech V_Θ jest przestrzenią parametru Θ , tzn. $\Theta \in V_\Theta$. Wtedy logarytm funkcji wiarygodności $\ln P : V_\Theta \rightarrow \mathbb{R}$ jest funkcją określoną na przestrzeni V_Θ o wartościach w zbiorze liczb rzeczywistych \mathbb{R} . Niech $\tilde{\Theta} \equiv (\tilde{\theta}_n)_{n=1}^N \in V_\Theta$ jest inną wartością parametru lub wartością estymatora $\hat{\Theta}$ parametru Θ . Rozwińmy w punkcie $\tilde{\Theta}$ funkcję $\ln P(\tilde{\Theta})$ w szereg Taylora wokół prawdziwej wartości Θ :

$$\ln \frac{P(\tilde{\Theta})}{P(\Theta)} = \sum_{n=1}^N \frac{\partial \ln P(\Theta)}{\partial \theta_n} (\tilde{\theta}_n - \theta_n) + \frac{1}{2} \sum_{n,n'=1}^N \frac{\partial^2 \ln P(\Theta)}{\partial \theta_{n'} \partial \theta_n} (\tilde{\theta}_n - \theta_n)(\tilde{\theta}_{n'} - \theta_{n'}) + R_3, \quad (3.1)$$

gdzie użyto oznaczenia $\frac{\partial P(\Theta)}{\partial \theta_n} \equiv \frac{\partial P(\tilde{\Theta})}{\partial \theta_n} \big|_{\tilde{\Theta}=\Theta}$, oraz podobnie dla wyższych rzędów rozwinięcia, a R_3 jest resztą rozwinięcia trzeciego rzędu.

Znaczenie zasady obserwowanej: Wszystkie człony w (3.1) są statystykami na przestrzeni próby \mathcal{B} , więc tak jak i układ równań wiarygodności, równanie (3.1) jest określone na poziomie obserwowanym. Jest ono żądaniem analityczności (logarytmu) funkcji wiarygodności na przestrzeni statystycznej \mathcal{S} , co stanowi punkt wyjścia dla konstrukcji obserwowanej, różniczkowej, strukturalnej zasady informacyjnej (3.11), słusznej niezależnie od wyprowadzonej w (3.12) postaci całkowej. Postać obserwowana wraz z zasadą wariacyjną (3.41) jest, obok postaci oczekiwanej (3.12), podstawą estymacji EFI równań ruchu teorii pola, lub równań generujących rozkład.

Zdefiniujmy obserwowaną strukturę układu \mathfrak{tF} w następujący sposób:

$$\mathfrak{tF} \equiv \ln \frac{P(\tilde{\Theta})}{P(\Theta)} - R_3. \quad (3.2)$$

Na poziomie obserwowanym (nazywanym czasami mikroskopowym) możemy rozwinięcie Taylora (3.1) zapisać następująco:

$$\Delta_{LHS} \equiv \sum_{n=1}^N 2 \frac{\partial \ln P}{\partial \theta_n} (\tilde{\theta}_n - \theta_n) - \sum_{n=1}^N 2 \frac{\mathfrak{tF}}{N} = \sum_{n,n'=1}^N \mathfrak{iF}_{nn'} (\tilde{\theta}_n - \theta_n)(\tilde{\theta}_{n'} - \theta_{n'}) \equiv \Delta_{RHS}, \quad (3.3)$$

gdzie \mathfrak{iF} jest znaną już z (2.3)-(2.4) obserwowaną macierzą informacyjną Fishera:

$$\mathfrak{iF} \equiv \left(-\frac{\partial^2 \ln P(\Theta)}{\partial \theta_{n'} \partial \theta_n} \right) = \left(-\frac{\partial^2 \ln P(\tilde{\Theta})}{\partial \tilde{\theta}_{n'} \partial \tilde{\theta}_n} \right) \bigg|_{\tilde{\Theta}=\Theta}, \quad (3.4)$$

która jako macierz odwrotna do macierzy kowariancji, jest symetryczna i dodatnio określona¹ (Rozdział 2.1.1). Oznacza to, że istnieje ortogonalna macierz U taka, że Δ_{RHS} występujące w (3.3), a zatem również Δ_{LHS} , może być zapisane w tzw. postaci normalnej [32]:

$$\Delta_{LHS} = \sum_{n=1}^N m_n \tilde{v}_n^2 = \sum_{n,n'=1}^N \mathfrak{iF}_{nn'} (\tilde{\theta}_n - \theta_n)(\tilde{\theta}_{n'} - \theta_{n'}) \equiv \Delta_{RHS}, \quad (3.5)$$

¹Co oznacza, że zakładamy, że funkcja $\ln P$ jest wypukła w otoczeniu prawdziwej wartości parametru Θ .

gdzie \tilde{v}_n są pewnymi funkcjami $\tilde{\theta}_n$, a m_n są elementami dodatnio określonej macierzy \mathfrak{mF} (otrzymanymi dla Δ_{LHS}), która z powodu równości (3.5) musi być równa macierzy diagonalnej otrzymanej dla Δ_{RHS} , tzn.:

$$\mathfrak{mF} = D^T U^T \mathfrak{iF} U D . \quad (3.6)$$

Macierz D jest diagonalną macierzą skalującą o elementach $d_n \equiv \sqrt{\frac{m_n}{\lambda_n}}$, gdzie λ_n są wartościami własnymi macierzy \mathfrak{iF} .

Związek (3.6) można zapisać w postaci ważnego strukturalnego równania macierzowego będącego bezpośrednią konsekwencją analityczności logarytmu funkcji wiarygodności na przestrzeni statystycznej \mathcal{S} oraz postaci normalnej formy kwadratowej (3.5):

$$\mathfrak{qF} + \mathfrak{iF} = 0 , \quad (3.7)$$

gdzie

$$\mathfrak{qF} = -U (D^T)^{-1} \mathfrak{mF} D^{-1} U^T , \quad (3.8)$$

nazwijmy *obserwowaną macierzą struktury*.

Dwa proste przypadki \mathfrak{qF} : Istnieją dwa szczególne przypadki, które prowadzą do prostych realizacji fizycznych.

Pierwszy z nich związany jest z założeniem, że rozkład jest *reguralny* [13]. Wtedy, zakładając dodatkowo, że dla wszystkich $n = 1, \dots, N$ zachodzi $\frac{\partial \ln P}{\partial \theta_n} = 0$, z równania (3.3) widzimy, że:

$$\mathfrak{mF} = (2 \delta_{nn'}) , \quad \tilde{v}_n = \sqrt{\frac{\mathfrak{tF}}{N}} \quad \text{oraz} \quad d_n = \sqrt{2/\lambda_n} . \quad (3.9)$$

Natomiast drugi przypadek związany jest z założeniem, że $\mathfrak{tF} = 0$ i wtedy z (3.3) otrzymuje się postać “równania master” (porównaj dalej (3.68)). W przypadku tym:

$$\mathfrak{mF} = \text{diag} \left(2 \frac{\partial \ln P}{\partial \theta_n} \right) , \quad \tilde{v}_n = \sqrt{\tilde{\theta}_n - \theta_n} , \quad \mathfrak{tF} = 0 \quad \text{oraz} \quad d_n = \sqrt{2 \frac{\partial \ln P}{\partial \theta_n} / \lambda_n} , \quad (3.10)$$

co oznacza, że nie istnieje złożona struktura układu.

Obserwowana strukturalna zasada informacyjna: Sumując wszystkie elementy zarówno obserwowanej macierzy informacyjnej Fishera \mathfrak{iF} jak i obserwowanej macierzy struktury \mathfrak{qF} , równanie macierzowe (3.7) prowadzi do *obserwowanej strukturalnej zasady informacyjnej* Frieden’a:

$$\sum_{n,n'=1}^N (\mathfrak{iF})_{nn'} + \sum_{n,n'=1}^N (\mathfrak{qF})_{nn'} = 0 . \quad (3.11)$$

Znaczenie analityczności P oraz postaci \mathfrak{iF} dla EFI: Analityczność funkcji wiarygodności na przestrzeni statystycznej \mathcal{S} , wyrażona istnieniem rozwinięcia w szereg Taylora (3.1) oraz symetryczność i dodatnia określoność obserwowanej macierzy informacyjnej Fishera (3.4) są zasadniczymi warunkami, które czynią analizę Friedena-Soffera w ogóle możliwą. Okazuje się jednak, że w ogólności, dla otrzymania równań EFI należy odwołać się dodatkowo do wprowadzonej poniżej całkowitej zasady strukturalnej.

Całkowa strukturalna zasada informacyjna: Całkując obie strony równania (3.11) po całej przestrzeni próby \mathcal{B} (lub na jej podprzestrzeni) z miarą $dy P(\Theta)$, gdzie jak zwykle stosujemy oznaczenie $dy \equiv d^N y$, otrzymujemy całkową postać *informacyjnej zasady strukturalnej*:

$$Q + I = 0, \quad (3.12)$$

gdzie I jest uogólnieniem pojemności informacyjnej Fishera (2.210) (por. (2.31)):

$$I = \int_{\mathcal{B}} dy P(\Theta) \sum_{n,n'=1}^N (\mathfrak{IF})_{nn'}, \quad (3.13)$$

natomiast Q jest informacją strukturalną (SI):

$$Q = \int_{\mathcal{B}} dy P(\Theta) \sum_{n,n'=1}^N (\mathfrak{QF})_{nn'}. \quad (3.14)$$

Pierwotnie, w innej, informatycznej formie i interpretacji, zasada (3.12) została zapostulowana w [8]. Powyższa, fizyczna postać zasady strukturalnej (3.12) została zapostulowana w [9], a następnie wyprowadzona, jak to przedstawiono powyżej w [10].

Obserwowana zasada informacyjna (3.11) jest równaniem strukturalnym współczesnych modeli fizycznych wyprowadzanych metodą EFI. Natomiast użyteczność oczekiwanej strukturalnej zasady informacyjnej (3.12) okaże się być jasna przy, po pierwsze określeniu zmodyfikowanej obserwowanej zasady strukturalnej, po drugie, przy definicji całkowitej fizycznej informacji (3.31) oraz po trzecie, przy sformułowaniu informacyjnej zasady wariacyjnej (3.41). Oczekiwana zasada strukturalna jako taka, tzn. w postaci całkowitej (3.12), nie jest rozwiązywana jednocześnie z zasadą wariacyjną, co jest czasami jej przypisywane.

3.1.1.1 Całka rozwinięcia Taylora

Scałkujemy (3.1) na całej przestrzeni próby \mathcal{B} (lub na jej podprzestrzeni) z miarą $dy P(\Theta)$. W wyniku otrzymujemy pewną całkową formę strukturalnego równania estymacji modeli:

$$\begin{aligned} & \int_{\mathcal{B}} dy P(\Theta) \left(\ln \frac{P(\tilde{\Theta})}{P(\Theta)} - R_3 - \sum_{n=1}^N \frac{\partial \ln P(\Theta)}{\partial \theta_n} (\tilde{\theta}_n - \theta_n) \right) \\ &= \frac{1}{2} \int_{\mathcal{B}} dy P(\Theta) \sum_{n,n'=1}^N \frac{\partial^2 \ln P(\Theta)}{\partial \theta_{n'} \partial \theta_n} (\tilde{\theta}_n - \theta_n) (\tilde{\theta}_{n'} - \theta_{n'}) . \end{aligned} \quad (3.15)$$

Wyrażenie po lewej stronie (3.15) ma postać zmodyfikowanej entropii względnej.

Następnie, definiując \tilde{Q} jako:

$$\tilde{Q} = \int_{\mathcal{B}} dy P(\Theta) \left(\mathfrak{TF} - \sum_{n=1}^N \frac{\partial \ln P}{\partial \theta_n} (\tilde{\theta}_n - \theta_n) \right) \quad (3.16)$$

otrzymujemy równanie będące całkową formą strukturalnej zasady informacyjnej:

$$-\tilde{Q} = \tilde{I} \equiv \frac{1}{2} \int_{\mathcal{B}} dy P(\Theta) \sum_{n,n'=1}^N \left(-\frac{\partial^2 \ln P}{\partial \theta_{n'} \partial \theta_n} \right) (\tilde{\theta}_n - \theta_n) (\tilde{\theta}_{n'} - \theta_{n'}) . \quad (3.17)$$

Uwaga: Równanie (3.17) jest wtórne wobec bardziej fundamentalnego równania strukturalnego (3.11) słusznego na poziomie obserwowanym, tzn. pod całą. Chociaż równanie (3.17) nie jest bezpośrednio wykorzystywane w metodzie EFI, to jest ono stosowane do badania własności nieobciążonych estymatorów $\tilde{\Theta}$ parametrów Θ [24]. Zagadnienie to wykracza poza zakres skryptu.

3.1.1.2 I oraz Q dla parami niezależnych zmiennych położeniowych próby

Rozważmy jeszcze postać SI wyrażoną w amplitudach w szczególnym przypadku zmiennych Y_n parami niezależnych. W takim przypadku amplituda q_n nie zależy od \mathbf{y}_n dla $n' \neq n$, czyli ma postać $q_n(\mathbf{y}_n)$, natomiast (\mathbf{iF}) jest diagonalna, tzn. ma postać:

$$(\mathbf{iF})_{nn'} = \delta_{nn'} \mathbf{iF}_{nn} \equiv \mathbf{iF}_n , \quad (3.18)$$

gdzie $\delta_{nn'}$ jest deltą Kroneckera. W takim razie, zgodnie z (3.6) oraz (3.8) obserwowana macierz strukturalna jest diagonalna i jej ogólna postać jest następująca:

$$(\mathbf{\Phi})_{nn'} = \delta_{nn'} \mathbf{\Phi}_{nn} \left(q_n(\mathbf{y}_n), q_n^{(r)}(\mathbf{y}_n) \right) \equiv \mathbf{\Phi}_n(q_n(\mathbf{y}_n)) , \quad (3.19)$$

tzn. nie zależy od amplitud $q_{n'}(\mathbf{y}_{n'})$ i jej pochodnych dla $n' \neq n$. Powyżej $q_n^{(r)}(\mathbf{y}_n)$ oznaczają pochodne rzędu $r = 1, 2, \dots$. Zobaczymy, że dla teorii pola w $\mathbf{\Phi}_n$ pojawiają się pochodne co najwyżej pierwszego rzędu. Fakt ten wynika stąd, że swobodne pola rangi N , z którymi będziemy mieli do czynienia, będą spełniały równanie Kleina-Gordona.

Uwaga: Oznaczenie $\mathbf{\Phi}_n$, jak również jawne zaznaczenie w argumencie obserwowanej SI tylko amplitudy $q_n(\mathbf{y}_n)$, będą stosowane w dalszej części skryptu.

Wykorzystując (3.18) oraz (3.19), pojemność informacyjna (3.13) przyjmuje w rozważanym przypadku postać:

$$I = \int_{\mathcal{B}} dy \, i = \int_{\mathcal{B}} dy \, P(\Theta) \sum_{n=1}^N \mathbf{iF}_n , \quad (3.20)$$

natomiast informacja strukturalna (3.14) jest następująca:

$$Q = \int_{\mathcal{B}} dy \, q = \int_{\mathcal{B}} dy \, P(\Theta) \sum_{n=1}^N \mathbf{\Phi}_n(q_n(\mathbf{y}_n)) . \quad (3.21)$$

Powyżej i jest gęstością pojemności informacyjnej:

$$i := P(\Theta) \sum_{n=1}^N \mathbf{iF}_n , \quad (3.22)$$

natomiast q jest gęstością informacji strukturalnej:

$$q := P(\Theta) \sum_{n=1}^N \mathbf{\Phi}_n(q_n(\mathbf{y}_n)) . \quad (3.23)$$

Obserwowana zasada strukturalna zapisana w gęstościach: Zarówno i jak i q są określone na poziomie obserwowanym. Zatem korzystając z (3.18) oraz (3.19), możemy *obserwowaną* informacyjną zasadę strukturalną (3.11) zapisać w postaci:

$$i + q = 0 . \quad (3.24)$$

Zasada ta, a raczej jej zmodyfikowana wersja, jest obok wariacyjnej zasady informacyjnej, wykorzystywana w celu otrzymania równań ruchu (bądź równań generujących rozkład) metody EFI. Zarówno zmodyfikowana obserwowana zasada strukturalna jak i zasada wariacyjna są określone poniżej.

Uwaga: W treści skryptu gęstość pojemności informacyjnej i jest zawsze związana z postacią (3.4) obserwowanej informacji Fishera i_F .

Na koniec zauważmy, że ze względu na unormowanie rozkładów brzegowych $\int d^4\mathbf{y}_n p_n(\mathbf{y}_n|\theta_n) = 1$, postać Q podaną w (3.21) można zapisać następująco:

$$Q = \sum_{n=1}^N \int d^4\mathbf{y}_n p_n(\mathbf{y}_n|\theta_n) \mathfrak{F}_n(q_n(\mathbf{y}_n)) . \quad (3.25)$$

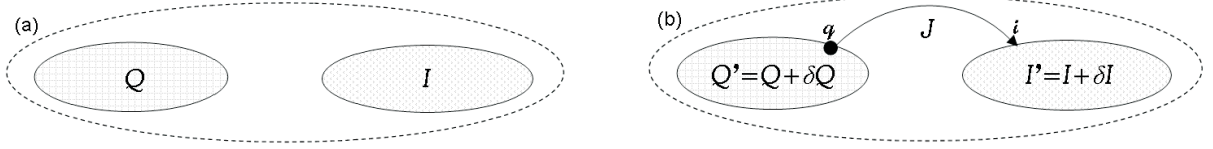
Ważna kinematyczna postać I zostanie wprowadzona w Rozdziale 3.3, natomiast postacie Q będą pojawiały się w toku rozwiązywania konkretnych fizycznych problemów.

3.2 Przepływ informacji

Informacja Fishera I_F jest infinitezymalnym typem entropii Kulback-Leibler’a (Rozdział 2.6.1) wzór (2.201). W statystycznej estymacji KL służy jako narzędzie analizy wyboru modelu [23, 33], o czym możemy się przekonać, zauważając, że jest ona związana z wartością oczekiwaną statystyki ilorazu wiarygodności (1.58), wprowadzonej w Rozdziale 1.3.2, właśnie w celu porównywania wiarygodności modeli. Chociażby z tego powodu, pojawia się przypuszczenie, że pojemność informacyjna I mogłaby, po nałożeniu, jak się okazuje strukturalnej i wariacyjnej zasady informacyjnej [8, 9], stać się podstawą równań ruchu (lub równań generujących rozkład) układu fizycznego. Równania te miałyby być najlepsze z punktu widzenia zasad informacyjnych, co jest sednem metody EFI Friedena-Soffera.

Zgodnie z Rozdziałem 2.8, główna statystyczna myśl stojąca za metodą EFI jest następująca: próbkowanie czasoprzestrzeni następuje przez sam układ nawet wtedy, gdy on sam nie jest poddany rzeczywistemu pomiarowi. Sprawę należałoby rozumieć tak, że układ dokonuje próbkowania czasoprzestrzeni używając charakterystycznego, swojego własnego pola (i związanej z nim amplitudy) rangi N , która jest wymiarem próby, próbując swoimi kinematycznymi “Fisherowskimi” stopniami swobody przestrzeni położeń jemu dostępną. Przejście od postaci statystycznej pojemności informacyjnej (2.220) do jej reprezentacji kinematycznej zostanie omówione poniżej w Rozdziale 3.3.

Rozważmy następujący, informacyjny schemat układu. Zanim nastąpi pomiar, którego dokonuje sam układ, ma on pojemność informacyjną I zawartą w swoich kinematycznych stopniach swobody oraz informację strukturalną Q układu zawartą w swoich strukturalnych stopniach swobody, jak to przedstawiono symbolicznie na poniższym Rysunku.



Rysunek 3.1: Panel: (a) Układ przed pomiarem : Q jest SI układu zawartą w strukturalnych stopniach swobody, a I jest pojemnością informacyjną układu zawartą w kinematycznych stopniach swobody. (b) Układ po pomiarze: Q' jest SI , a I' jest pojemnością informacyjną układu po pomiarze. Ponieważ transfer informacji (TI) w pomiarze przebiega z $J \geq 0$ zatem $\delta Q = Q' - Q \leq 0$ oraz $\delta I = I' - I \geq 0$. W pomiarze idealnym $\delta I = -\delta Q$.

“W chwili włączenia” pomiaru, podczas którego transfer informacji (TI) przebiega zgodnie z następującymi zasadami (Rysunek 3.1):

$$J \geq 0, \quad \text{zatem} \quad \delta I = I' - I \geq 0, \quad \delta Q = Q' - Q \leq 0, \quad (3.26)$$

gdzie I' , Q' są odpowiednio IF oraz SI układu po pomiarze, natomiast J jest dokonanym transferem informacji (TI).

Postulujemy, że w pomiarze TI “w punkcie q ” jest idealny, co oznacza, że:

$$Q = Q' + J = Q + \delta Q + J, \quad \text{zatem} \quad \delta Q = -J. \quad (3.27)$$

Oznacza to, że “w punkcie q ” przekazana jest cała zmiana SI .

Z drugiej strony “w punkcie i ” zasada związana z TI jest następująca:

$$I' \leq I + J \quad \text{zatem} \quad 0 \leq \delta I = I' - I \leq J. \quad (3.28)$$

Dlatego

$$\text{ponieważ} \quad J \geq 0, \quad \text{zatem} \quad |\delta I| \leq |\delta Q|, \quad (3.29)$$

co jest rozsądnym rezultatem, gdyż w pomiarze może nastąpić utrata informacji. Gdyby “w punkcie i ” TI był idealny, wtedy cały pomiar byłby idealny, tzn.:

$$\delta Q = -\delta I \Leftrightarrow \text{pomiar idealny}. \quad (3.30)$$

W [9, 16] zostało zapostulowane istnienie nieujemnej addytywnej całkowitej (totalnej) fizycznej informacji (TFI):

$$K = I + Q \geq 0. \quad (3.31)$$

Wybór intuicyjnego warunku $K \geq 0$ [16] jest związany ze *strukturalną zasadą informacyjną* zapisaną w postaci obserwowanej:

$$\sum_{n,n'=1}^N (\mathbf{iF})_{nn'} + \kappa \sum_{n,n'=1}^N (\mathbf{qF})_{nn'} = 0 \quad (3.32)$$

lub oczekiwanej:

$$I + \kappa Q = 0. \quad (3.33)$$

Dla szczególnego przypadku $\kappa = 1$, w Rozdziale 3.1.1 została wyprowadzona [10] postać obserwowana zasady strukturalnej (3.11):

$$\sum_{n,n'=1}^N (\mathfrak{IF})_{nn'} + \sum_{n,n'=1}^N (\mathfrak{QF})_{nn'} = 0 \quad \text{dla } \kappa = 1, \quad (3.34)$$

oraz jej oczekiwany odpowiednik (3.12):

$$I + Q = 0 \quad \text{dla } \kappa = 1. \quad (3.35)$$

Współczynnik κ został nazwany w [8] współczynnikiem efektywności. W praktyce przyjmuje on dwie możliwe wartości [8]:

$$\kappa = 1 \quad \vee \quad \frac{1}{2}. \quad (3.36)$$

Jego znaczenie zostanie omówione w Rozdziale 4. W przypadku określonym w (3.35), otrzymujemy całkowitą fizyczną informację K równą:

$$K = I + Q = 0 \quad \text{dla } \kappa = 1. \quad (3.37)$$

W końcu zauważmy, że w zgodzie z zapostulowanym zachowaniem się układu w pomiarze, otrzymaliśmy z warunków (3.27) i (3.28) nierówność $\delta I \leq J = -\delta Q$, z czego wynika, że:

$$K' = I' + Q' \leq (I + J) + (Q - J) = I + Q = K \Rightarrow K' \leq K. \quad (3.38)$$

Dla pomiaru idealnego (3.30) otrzymaliśmy $\delta I = -\delta Q$ skąd $K' = K$, co oznacza, że informacja fizyczna TFI pozostaje w tym przypadku niezmienną. Jeśli pomiar idealny byłby wykonany na poziomie próbkowania czasoprzestrzeni przez sam układ, wtedy warunek ten mógłby prowadzić do wariacyjnej zasady informacyjnej (3.41), tzn.:

$$\delta I = -\delta Q \Rightarrow \delta(I + Q) = 0. \quad (3.39)$$

Chociaż rozumowanie powyższe wydaje się być rozsądne, jednak ściśle mówiąc słuszność przyjęcia zasad informacyjnych, strukturalnej oraz wariacyjnej, powinno wynikać z dwóch rzeczy. Po pierwsze z ich wyprowadzenia, a po drugie z ich użyteczności. Wyprowadzenie zasady strukturalnej (dla $\kappa = 1$) zostało pokazane w Rozdziale 3.1.1.

Natomiast powyższe wnioskowanie, które doprowadziło do warunku (3.39) oraz sama implikacja wewnętrznie, może służyć jedynie jako przesłanka słuszności zasady wariacyjnej. W Rozdziale 2.2.3 stwierdziliśmy, że jej słuszność wynika z żądania aby rozkład empiryczny oraz rozkład wyestymowany metodą EFI, leżały na wspólnej geodezyjnej w przestrzeni statystycznej \mathcal{S} .

Co do użyteczności zasady wariacyjnej w metodzie EFI, to jest ona oczywista, bowiem prowadzi ona do owocnego w zastosowaniach równania Eulera-Lagrange'a.

Analityczny przypadek układu równań informacyjnych metody EFI: Obserwowana zasada strukturalna zapisana w gęstościach (3.24), ale uwzględniająca postać obserwowanej zasady strukturalnej z κ (3.32), jest następująca:

$$i + \kappa q = 0. \quad (3.40)$$

Drugą zasadą informacyjną jest *zasada wariacyjna* (skalarna). Ma ona postać [9]:

$$\delta K = \delta(I + Q) = 0 \Rightarrow K = I + Q \text{ jest ekstremalne}. \quad (3.41)$$

Warunek geometrycznej struktury na \mathcal{S} : Kolejnym warunkiem narzuconym na rozwiązania metody EFI, a oczywistym od początku analizy, jest warunek normalizacji i regularności rozkładu prawdopodobieństwa. Warunek ten oznacza możliwość przejścia, podanego w (2.31) i (2.36), od pierwotnej postaci obserwowanej informacji Fishera (3.4) do postaci potrzebnej dla zdefiniowania przestrzeni statystycznej \mathcal{S} jako przestrzeni metrycznej z metryką Rao-Fishera (2.37) i α -koneksją (2.52) (por. Rozdział 2.2). Obie postaci obserwowanej informacji Fishera, pierwsza \mathbf{iF} , (3.4), która jest pierwotną formą z punktu widzenia analityczności funkcji wiarygodności oraz druga, *metryczna* $\tilde{\mathbf{iF}}$, (2.5), która jest istotna dla geometrycznej analizy modelu, są równoważne tylko na poziomie oczekiwanym, tzn. pod całką (por. Rozdział 2.31). Powyższe rozważania zostaną zilustrowane przykładami zawartymi w dalszej części skryptu.

Podstawowy układ równań informacyjnych EFI i zmodyfikowane równanie strukturalne: Aby wyjaśnić powyższy problem na wstępnym, symbolicznym poziomie, wprowadźmy zmodyfikowane równanie strukturalne, uwzględniające również współczynnik κ występujący w (3.40). Niech $\tilde{\mathbf{iF}}$ jest kwadratową postacią obserwowanej informacji Fishera (2.5) tak, że odpowiadająca jej gęstość pojemności informacyjnej:

$$\tilde{i} := P(\Theta) \sum_{n,n'=1}^N \tilde{\mathbf{iF}}_{nn'} , \quad (3.42)$$

daje na poziomie oczekiwanym pojemność informacyjną $I = \int_{\mathcal{B}} dy \tilde{i} = \int_{\mathcal{B}} dy i$.

Wprowadźmy zamiast (3.40) *zmodyfikowaną obserwowaną zasadę strukturalną* zapisaną w następujący sposób:

$$\tilde{i}' + \tilde{\mathbf{C}} + \kappa q = 0 , \quad \text{przy czym} \quad I = \int_{\mathcal{B}} dy \tilde{i} = \int_{\mathcal{B}} dy (\tilde{i}' + \tilde{\mathbf{C}}) , \quad (3.43)$$

gdzie $\tilde{\mathbf{C}}$ jest pochodną zupełną, która wynika z całkowania przez części całki $I = \int_{\mathcal{B}} dy \tilde{i}$. Ponieważ $I = \int_{\mathcal{B}} dy \tilde{i}$ zatem informacyjna zasada wariacyjna ma postać:

$$\delta(I + Q) = \delta \int_{\mathcal{B}} dy (\tilde{i} + q) = 0 . \quad (3.44)$$

Rozwiązanie równań (3.43) oraz (3.44) jest równoważne rozwiązaniu równań (3.40) oraz (3.41) co najmniej pod całką, tzn.:

$$\int_{\mathcal{B}} dy (\tilde{i}' + \tilde{\mathbf{C}} + \kappa q) = 0 \quad \Leftrightarrow \quad \int_{\mathcal{B}} dy (i + \kappa q) = 0 . \quad (3.45)$$

Powyższą symboliczną konstrukcję (3.43) zaprezentujemy na przykładach w dalszej części skryptu (Rozdział 5). Jej zrozumienie jest następujące: Modele „nie do końca równoważne” pod względem analitycznym są, z dokładnością do wycałkowania I przez części, równoważne pod względem metrycznym. To znaczy, istnieje pewien związek pomiędzy ich różniczkowalnością, a mianowicie wszystkie one są metrycznie (a więc na poziomie całkowym) równoważne modelowi analitycznemu, tzn. posiadającemu rozwinięcie w szereg Taylora.

Podsumowanie. Należy podkreślić, że równanie całkowe (3.37) oraz zasada wariacyjna (3.41) *nie* tworzą pary równań metody EFI rozwiązywanych samospójnie.

Natomiast obie zasady, obserwowana zmodyfikowana zasada strukturalna (3.43) oraz zasada wariacyjna (3.44) są podstawą metody estymacyjnej EFI. Tworzą one układ dwóch równań różniczkowych dla wprowadzonych w Rozdziale 2.5 amplitud układu (2.168). Układ ten może być zgodny, dając samospójne rozwiązanie dla amplitud [8] i prowadząc przy $\kappa = 1$ lub $1/2$ do dobrze znanych modeli teorii pola (Rozdział 4) lub modeli fizyki statystycznej (Rozdział 5). Ponadto, strukturalna (wewnętrzna) zasada informacyjna (3.33) [9] jest operacyjnie równoważna zapostulowanej przez Frieden'a [8], więc jako wyprowadzona powinna mieć przynajmniej taką samą moc przewidywania jak i ona. Wiele z podstawowych modeli zostało już wyliczonych [8], jednak ich ponowne przeliczenie [16] przy powyższej podanej interpretacji informacji fizycznej K może dać lepsze zrozumienie samej metody EFI i jej związku z istniejącym już modelowaniem zjawisk w fizyce oraz jej ograniczeń.

Zasada ekwipartycji entropii względnej: W końcu, w strukturalnej zasadzie informacyjnej ciekawe jest również to, że stanowi ona warunek zerowego podziału dla TFI , który jest dawno poszukiwanym warunkiem zasady ekwipartycji entropii (w tym przypadku infinitezymalnej entropii względnej).

Uwaga o podejściu Friedena: Wspomnieliśmy o tym, że pomysł metody EFI pochodzi od Friedena. Jednak mówiąc w skrócie, Frieden i Soffer [8] podeszli inaczej do informacji strukturalnej. W [8] wprowadzono tzw. informację związaną J , która ma interpretację informacji zawartej w układzie przed pomiarem. Chociaż, aksjomaty Frieden'a są równoważne powyższym warunkom (3.33) oraz (3.41), o ile $J = -Q$, to jednakże różnica pomiędzy podejściami jest widoczna. A mianowicie, o ile w podejściu Friedena-Soffera układ doświadcza transferu informacji $J \rightarrow I$, mając w każdej chwili czasu tylko jeden z tych typów informacji, o tyle w naszym podejściu system jest charakteryzowany jednocześnie przez I oraz Q w każdej chwili czasu.

Uwaga o podobieństwie EFI i teorii Jaynes'a: Metoda EFI zaproponowana przez Friedena i Soffera [8] jest konsekwencją postulatu podobnego do zasady Jaynes'a². Mianowicie podobieństwo obu teorii leży w tym, że poprzez zasadę wariacyjną wiążą one strukturalne (Boltzmann'owskie) stopnie swobody z kinematycznymi (Shannona) stopniami swobody³.

Według podejścia Jaynes'a, maksymalizacja entropii Shannona względem prawdopodobieństw mikrostanu układu, posiadającego znane własności, np. ustaloną energię, umożliwia identyfikację termodynamicznej entropii Boltzmanna jako zmaksymalizowanej entropii Shannona, a następnie na konstrukcję funkcji stanu, np. energii swobodnej.

3.3 Kinetyczna postać informacji Fishera

Centralna część pracy Frieden'a i Soffer'a związana jest z transformacją postaci pojemności informacyjnej I zadanej równaniem (2.220) oraz (2.219) do tzw. postaci kinematycznej wykorzystywanej w teorii pola oraz fizyce statystycznej. W obecnym rozdziale zaprezentujemy podstawowe założenia, które doprowadziły do konstrukcji kinematycznego członu (całki) działania dla czterowymiarowych modeli teorii pola. Przejście to ma następującą postać [8].

²E. Jaynes, Information Theory and Statistical Mechanics, *Phys.Rev.* 106, 620–630 (1957). E. Jaynes, Information Theory and Statistical Mechanics. II, *Phys.Rev.* 108, 171–190 (1957).

³Metod tych nie należy jednak utożsamiać. Należy pamiętać, że macierz informacyjna Fishera, wykorzystywana w EFI, oraz entropia Shannona, wykorzystywana w podejściu Jaynes'a, są różnymi pojęciami. Macierz informacyjna Fishera jest Hessianem (2.171) entropii Shannona.

Zgodnie z podstawowym założeniem Friedena-Soffera, N -wymiarowa próbka $\mathbf{y}_n \equiv (\mathbf{y}_n^\nu)$ jest pobierana przez układ posiadający rozkład $p_n(\mathbf{y}_n)$, gdzie obok indeksu próby $n = 1, 2, \dots, N$ wprowadzono indeks (czaso)przestrzenny $\nu = (0), 1, 2, 3$. Zgodnie z Rozdziałem 2.5, wzór (2.168), metryka Fishera na (pod)rozmaitości \mathcal{S} prowadzi w naturalny sposób do pojęcia rzeczywistej amplitudy $q(\mathbf{y}_n|\theta_n) \equiv \sqrt{p(\mathbf{y}_n|\theta_n)}$ pola układu. Od razu skorzystano też z zapisu, który sugeruje niezależność rozkładu dla \mathbf{y}_n od θ_m , gdy $m \neq n$.

Jak przedstawiliśmy w Rozdziale 2.8 pojemność informacyjna (2.231) może zostać zapisana jako (2.235):

$$I \equiv I(\Theta) = 4 \sum_{n=1}^N \int_{\mathcal{Y}} d^4 \mathbf{y}_n \sum_{\nu=0}^3 \left(\frac{\partial q_n(\mathbf{y}_n|\theta_n)}{\partial \theta_{n\nu}} \frac{\partial q_n(\mathbf{y}_n|\theta_n)}{\partial \theta_n^\nu} \right). \quad (3.46)$$

Addytywny rozkład położeń i reguła łańcuchowa: Niech $\mathbf{x}_n \equiv (\mathbf{x}_n^\nu)$ są przesunięciami (np. addytywnymi fluktuacjami) danych $\mathbf{y}_n \equiv (\mathbf{y}_n^\nu)$ od ich wartości oczekiwanych θ_n^ν , tzn.:

$$\mathbf{y}_n^\nu = \theta_n^\nu + \mathbf{x}_n^\nu. \quad (3.47)$$

Przesunięcia \mathbf{x}_n^ν są zmiennymi Fisher'owskimi, spełniając warunek $\frac{\partial \mathbf{x}_n^\nu}{\partial \mathbf{x}^\mu} = \delta_\mu^\nu$, (2.223).

Odwołując się do "reguły łańcuchowej" dla pochodnej:

$$\frac{\partial}{\partial \theta_n^\nu} = \frac{\partial(\mathbf{y}_n^\nu - \theta_n^\nu)}{\partial \theta_n^\nu} \frac{\partial}{\partial(\mathbf{y}_n^\nu - \theta_n^\nu)} = - \frac{\partial}{\partial(\mathbf{y}_n^\nu - \theta_n^\nu)} = - \frac{\partial}{\partial \mathbf{x}_n^\nu} \quad (3.48)$$

oraz uwzględniając $d^4 \mathbf{x}_n = d^4 \mathbf{y}_n$, co wynika z tego, że parametry θ_n są stałymi, możemy przejść od postaci statystycznej (3.46) do postaci kinematycznej IF :

$$I = 4 \sum_{n=1}^N \int_{\mathcal{X}_n} d^4 \mathbf{x}_n \sum_{\nu} \frac{\partial q_n(\mathbf{x}_n)}{\partial \mathbf{x}_{n\nu}} \frac{\partial q_n(\mathbf{x}_n)}{\partial \mathbf{x}_n^\nu}, \quad (3.49)$$

gdzie $d^4 \mathbf{x}_n = d\mathbf{x}_n^0 d\mathbf{x}_n^1 d\mathbf{x}_n^2 d\mathbf{x}_n^3$. W (3.49) wprowadzono oznaczenie:

$$q_n(\mathbf{x}_n) \equiv q_n(\mathbf{x}_n + \theta_n|\theta_n) = q_n(\mathbf{y}_n|\theta_n), \quad (3.50)$$

pozostawiając całą informację o θ_n w indeksie n amplitudy $q_n(\mathbf{x}_n)$.

Kinematyczna postać IF dla q_n : Zakładając, że zakres zmienności wszystkich \mathbf{x}_n^ν jest dla każdego n taki sam, możemy pominąć indeks n przy tej zmiennej (ale nie przy amplitudzie q_n), otrzymując postać:

$$I = 4 \sum_{n=1}^N \int_{\mathcal{X}} d^4 \mathbf{x} \sum_{\nu} \frac{\partial q_n(\mathbf{x})}{\partial \mathbf{x}_\nu} \frac{\partial q_n(\mathbf{x})}{\partial \mathbf{x}^\nu}, \quad (3.51)$$

którą wykorzystamy przy wyprowadzeniu równań generujących fizyki statystycznej [8], ale która została również wykorzystana do wyprowadzenia elektrodynamiki Maxwella metodą EFI [8].

Uwaga: Wymiar próby N jest rangą pola układu zdefiniowanego jako zbiór amplitud $(q_n(\mathbf{x}_n))_{n=1}^N$.

W Rozdziale 3.1.1 pokazaliśmy, że strukturalna zasada informacyjna $I + Q = 0$ jest artefaktem istnienia rozwinięcia $\ln P(\tilde{\Theta})$ w szereg Taylora⁴ wokół prawdziwej wartości parametru Θ . Obecnie znamy już ogólną postać kinematyczną I części pomiarowej zasady strukturalnej. W metodzie EFI, jej część strukturalna Q ma postać zależną od np. fizycznych więzów nałożonych na układ. Zagadnieniem tym zajmujemy się w kolejnych Rozdziałach 4 oraz 5.

Amplitudy zespolone: Kolejnym założeniem jest konstrukcja *składowych funkcji falowej* składanych z amplitud⁵ w następujący sposób [8]:

$$\psi_n(\mathbf{x}_{2n-1}, \mathbf{x}_{2n}) \equiv \frac{1}{\sqrt{N}} (q_{2n-1}(\mathbf{x}_{2n-1}) + i q_{2n}(\mathbf{x}_{2n})) , \quad n = 1, \dots, N/2 . \quad (3.52)$$

Powyższa postać jest uogólnieniem konstrukcji Friedena, który tworząc funkcję falową układu złożył n -tą *składową funkcji falowej* z amplitud w następujący sposób [8]:

$$\psi_n(\mathbf{x}) \equiv \frac{1}{\sqrt{N}} (q_{2n-1}(\mathbf{x}) + i q_{2n}(\mathbf{x})) , \quad n = 1, \dots, N/2 . \quad (3.53)$$

Dokładniej mówiąc, aby posłużenie się funkcją falową (3.53) miało sens, musi przynajmniej pod całką zachodzić równoważność zmiennych:

$$\mathbf{x}_n \equiv \mathbf{x} \quad \text{dla wszystkich } n = 1, 2, \dots, N . \quad (3.54)$$

Założenie to całkiem wystarcza przy liczeniu wartości oczekiwanych oraz prawdopodobieństw.

Przy założeniu postaci (3.53) dla n -tej składowej⁶, postać *funkcji falowej* Friedena jest następująca:

$$\psi(\mathbf{x}) \equiv (\psi_n(\mathbf{x}))_{n=1}^{N/2} . \quad (3.55)$$

Zbiór $N/2$ składowych funkcji falowych ψ_n nazwijmy *funkcją falową układu rangi N* .

Zauważmy, że zachodzą następujące równości:

$$\sum_{n=1}^N q_n^2 = (q_1^2 + q_3^2 + \dots + q_{N-1}^2) + (q_2^2 + q_4^2 + \dots + q_N^2) = \sum_{n=1}^{N/2} (q_{2n-1})^2 + (q_{2n})^2 \quad (3.56)$$

i analogicznie:

$$\sum_{n=1}^N \frac{\partial q_n}{\partial \mathbf{x}_{n\nu}} \frac{\partial q_n}{\partial \mathbf{x}_n^\nu} = \sum_{n=1}^{N/2} \left(\frac{\partial q_{2n-1}}{\partial \mathbf{x}_{n\nu}} \frac{\partial q_{2n-1}}{\partial \mathbf{x}_n^\nu} + \frac{\partial q_{2n}}{\partial \mathbf{x}_{n\nu}} \frac{\partial q_{2n}}{\partial \mathbf{x}_n^\nu} \right) . \quad (3.57)$$

Zakładając dla wszystkich poniższych rozważań słuszność (3.54), przynajmniej pod całką, oraz postać funkcji falowych (3.53), dokonajmy następującego ciągu przekształceń dla (3.51):

$$\begin{aligned} I &= 4 \sum_{n=1}^N \int_{\mathcal{X}} d^4 \mathbf{x} \sum_{\nu} \frac{\partial q_n}{\partial \mathbf{x}_{\nu}} \frac{\partial q_n}{\partial \mathbf{x}^{\nu}} = 4 \sum_{n=1}^{N/2} \int_{\mathcal{X}} d^4 \mathbf{x} \sum_{\nu} \left[\frac{\partial q_{2n-1}}{\partial \mathbf{x}_{\nu}} \frac{\partial q_{2n-1}}{\partial \mathbf{x}^{\nu}} + \frac{\partial q_{2n}}{\partial \mathbf{x}_{\nu}} \frac{\partial q_{2n}}{\partial \mathbf{x}^{\nu}} \right] \\ &= 4N \sum_{n=1}^{N/2} \int_{\mathcal{X}} d^4 \mathbf{x} \sum_{\nu} \frac{1}{\sqrt{N}} \frac{\partial (q_{2n-1} - i q_{2n})}{\partial \mathbf{x}_{\nu}} \frac{1}{\sqrt{N}} \frac{\partial (q_{2n-1} + i q_{2n})}{\partial \mathbf{x}^{\nu}} , \end{aligned} \quad (3.58)$$

⁴W [10] była użyta miara $d^N \mathbf{x} P(\Theta)$ zamiast $d^N \mathbf{y} P(\Theta)$. Nie zmienia to jednak dowodu strukturalnej zasady informacyjnej, lecz poszerza jego zastosowanie na sytuacje, które *nie posiadają niezmienniczości przesunięcia*, założenia nie wykorzystywano w dowodzie.

⁵Amplitudy q_n są w przypadku rozkładów ciągłych związane z p_n , które są *gęstościami* prawdopodobieństw.

⁶O ile nie będzie to prowadziło do nieporozumień, będziemy pominiali słowo “składowa”.

gdzie w ostatnim przejściu skorzystano z przekształcenia typu:

$$q_{2n-1}^k q_{2n-1}^k + q_{2n}^k q_{2n}^k = (q_{2n-1}^k - i q_{2n}^k)(q_{2n-1}^k + i q_{2n}^k) , \quad (3.59)$$

z indeksem k oznaczającym pochodną rzędu $k = 0, 1, \dots$.

Kinematyczna postać IF dla ψ : Odwołując się do definicji (3.53) funkcji falowej, otrzymujemy:

$$I = 4N \sum_{n=1}^{N/2} \int_{\mathcal{X}} d^4 \mathbf{x} \sum_{\nu} \frac{\partial \psi_n^*(\mathbf{x})}{\partial \mathbf{x}_{\nu}} \frac{\partial \psi_n(\mathbf{x})}{\partial \mathbf{x}^{\nu}} . \quad (3.60)$$

Pojemność informacyjna (3.60) ma typową postać np. dla relatywistycznej mechaniki falowej, odpowiadającą części kinetycznej całki działania. Dlatego właśnie oczekiwaną informację Fishera nazwał Frieden *informacją kinetyczną*. W [8] użyto jej do wyprowadzenia równań Kleina-Gordona oraz Diraca metodą EFI [8].

Rozkładu prawdopodobieństwa przesunięcia w układzie: Korzystając z twierdzenia o prawdopodobieństwie całkowitym, gęstość rozkładu prawdopodobieństwa przesunięcia (lub fluktuacji) w układzie może być zapisana następująco [8]:

$$\begin{aligned} p(\mathbf{x}) &= \sum_{n=1}^N p(\mathbf{x}|\theta_n) r(\theta_n) = \sum_{n=1}^N p_n(\mathbf{x}_n|\theta_n) r(\theta_n) = \frac{1}{N} \sum_{n=1}^N q_n^2(\mathbf{x}_n|\theta_n) \\ &= \frac{1}{N} \sum_{n=1}^N q_n^2(\mathbf{x}) , \end{aligned} \quad (3.61)$$

gdzie skorzystano z założenia, że n -ta wartość oczekiwana θ_n nie ma dla $m \neq n$ wpływu na rozkład przesunięcia \mathbf{x}_m oraz jak zwykle z postaci amplitudy $q_n^2 = p_n$. Prawdopodobieństwo p_n jest prawdopodobieństwem pojawienia się wartości \mathbf{x}_n zmiennej losowej przesunięcia (lub fluktuacji) z rozkładu generowanego z parametrem θ_n , tzn. ma ono interpretację prawdopodobieństwa warunkowego $p_n(\mathbf{x}_n|\theta_n)$. Funkcję $r(\theta_n) = \frac{1}{N}$ można nazwać funkcją “niewiedzy”, gdyż jej postać jest odzwierciedleniem całkowitego braku wiedzy odnośnie tego, która z N możliwych wartości θ_n pojawi się w konkretnym n -tym z N eksperymentów próby.

Postać rozkładu dla ψ : W końcu, korzystając z (3.53), (3.57), (3.59) oraz (3.61) uidać, że:

$$p(\mathbf{x}) = \sum_{n=1}^{N/2} \psi_n^*(\mathbf{x}) \psi_n(\mathbf{x}) \quad (3.62)$$

jest gęstością rozkładu prawdopodobieństwa przesunięcia (lub fluktuacji) \mathbf{x} w układzie opisanym funkcją falową (3.55).

Uwaga o różnicy z podejściem Friedena: W całym powyższym wyprowadzeniu nie użyliśmy podstawowego założenia Friedena-Soffera o *niezmienniczości rozkładu ze względu na przesunięcie*, tzn.:

$$p_n(\mathbf{x}_n) = p_{x_n}(\mathbf{x}_n|\theta_n) = p_n(\mathbf{y}_n|\theta_n) , \quad \text{gdzie} \quad \mathbf{x}_n^{\nu} \equiv \mathbf{y}_n^{\nu} - \theta_n^{\nu} , \quad (3.63)$$

gdzie $\mathbf{y}_n \equiv (\mathbf{y}_n^{\nu})$, $\mathbf{x}_n \equiv (\mathbf{x}_n^{\nu})$ oraz $\theta_n \equiv (\theta_n^{\nu})$. Założenie to nie jest potrzebne przy wyprowadzeniu postaci (3.49) pojemności informacyjnej.

Uwaga: Co więcej, informacja o θ_n musi pozostać w rozkładzie p_n oraz jego amplitudzie q_n . Wcześniej umówiliśmy się, że indeks n zawiera tę informację. Po umiejscowieniu informacji o θ_n w indeksie n można, w razie potrzeby wynikającej np. z fizyki zjawiska, zażądać dodatkowo niezmienniczości ze względu na przesunięcie.

3.3.1 Postać kinematyczna pojemności zapisana w prawdopodobieństwie

Poniżej podamy postać kinematyczną pojemności zapisaną w (punktowych) prawdopodobieństwach próby. Postać ta jest bardziej pierwotna niż (3.46), chociaż w treści skryptu wykorzystywana jedynie w Dodatku.

Puntem wyjścia jest pojemność (2.233):

$$I \equiv I(\Theta) = \sum_{n=1}^N \int_{\mathcal{Y}} d^4 \mathbf{y}_n \frac{1}{p_n(\mathbf{y}_n|\theta_n)} \sum_{\nu=0}^3 \left(\frac{\partial p_n(\mathbf{y}_n|\theta_n)}{\partial \theta_{n\nu}} \frac{\partial p_n(\mathbf{y}_n|\theta_n)}{\partial \theta_n^\nu} \right).$$

Korzystając z przejścia do addytywnych przesunięć $\mathbf{x}_n \equiv (\mathbf{x}_n^\nu)$, (3.47), oraz z “reguły łańcuchowej” (3.48) dla pochodnej, otrzymujemy (podobnie do (3.49)) następującą *kinematyczną postać pojemności informacyjnej*, wyrażoną w prawdopodobieństwach:

$$I = \sum_{n=1}^N \int_{\mathcal{X}_n} d^4 \mathbf{x}_n \frac{1}{p_n(\mathbf{x}_n)} \sum_{\nu=0}^3 \left(\frac{\partial p_n(\mathbf{x}_n)}{\partial \mathbf{x}_{n\nu}} \frac{\partial p_n(\mathbf{x}_n)}{\partial \mathbf{x}_n^\nu} \right), \quad (3.64)$$

gdzie, podobnie jak poprzednio dla amplitud, pozostawiliśmy całą informację o θ_n w indeksie n rozkładu $p_n(\mathbf{x}_n)$.

Postać kinematyczna I zapisana w prawdopodobieństwie: W końcu, zakładając, że zakres zmienności wszystkich \mathbf{x}_n^ν jest dla każdego n taki sam, pomijamy indeks n przy tej zmiennej (ale nie przy rozkładzie p_n), otrzymując:

$$I = \sum_{n=1}^N \int_{\mathcal{X}} d^4 \mathbf{x} \frac{1}{p_n(\mathbf{x})} \sum_{\nu=0}^3 \left(\frac{\partial p_n(\mathbf{x})}{\partial \mathbf{x}_\nu} \frac{\partial p_n(\mathbf{x})}{\partial \mathbf{x}^\nu} \right). \quad (3.65)$$

Postać tą wykorzystamy w Dodatku jako pierwotną przy wyprowadzeniu elektrodynamiki Maxwella, granicy słabego pola w teorii grawitacji oraz twierdzenia I fizyki statystycznej.

3.4 Równania master

Podejdźmy nieco inaczej niż w Rozdziale 3.1.1 do problemu estymacji. Rozwińmy $P(\tilde{\Theta})$ w szereg Taylora wokół prawdziwej wartości parametru Θ i wycałkujmy po całej przestrzeni próby \mathcal{B} , otrzymując:

$$\begin{aligned} \int_{\mathcal{B}} dy \left(P(\tilde{\Theta}) - P(\Theta) \right) &= \int_{\mathcal{B}} dy \left(\sum_{n=1}^N \frac{\partial P(\Theta)}{\partial \theta_n} (\tilde{\theta}_n - \theta_n) \right. \\ &+ \left. \frac{1}{2} \sum_{n,n'=1}^N \frac{\partial^2 P(\Theta)}{\partial \theta_{n'} \partial \theta_n} (\tilde{\theta}_n - \theta_n) (\tilde{\theta}_{n'} - \theta_{n'}) + \dots \right), \end{aligned} \quad (3.66)$$

gdzie użyto oznaczenia $\frac{\partial P(\Theta)}{\partial \theta_n} \equiv \frac{\partial P(\tilde{\Theta})}{\partial \tilde{\theta}_n} \big|_{\tilde{\Theta}=\Theta}$ oraz podobnie dla wyższych rzędów rozwinięcia. Ponieważ całkowanie zostaje wykonane po całej przestrzeni próby \mathcal{B} , zatem biorąc pod uwagę warunek normalizacji $\int_{\mathcal{B}} dy P(\Theta) = \int_{\mathcal{B}} dy P(\tilde{\Theta}) = 1$ widzimy, że lewa strona równania (3.66) jest równa zero. Pomijając człony wyższego rzędu⁷, otrzymujemy:

$$\int_{\mathcal{B}} dy \left(\sum_{n=1}^N \frac{\partial P(\Theta)}{\partial \theta_n} (\tilde{\theta}_n - \theta_n) + \frac{1}{2} \sum_{n,n'=1}^N \frac{\partial^2 P(\Theta)}{\partial \theta_{n'} \partial \theta_n} (\tilde{\theta}_n - \theta_n) (\tilde{\theta}_{n'} - \theta_{n'}) \right) = 0. \quad (3.67)$$

Dla estymatorów $\tilde{\Theta}$ lokalnie nieobciążonych [6] równanie (3.67) przyjmuje dla konkretnych n oraz n' następującą postać **równania master**:

$$\int_{\mathcal{B}} dy \frac{\partial^2 P(\Theta)}{\partial \theta_{n'} \partial \theta_n} (\tilde{\theta}_n - \theta_n) (\tilde{\theta}_{n'} - \theta_{n'}) = 0, \quad n, n' = 1, 2, \dots, N. \quad (3.68)$$

Gdy parametr θ_n^ν ma index Minkowskiego ν , wtedy można pokazać, że wykorzystując $P = \prod_{n=1}^N p_n(\mathbf{y}_n)$ w (3.68) otrzymujemy, po przejściu do zmiennych Fisherowskich (porównaj (2.223)), równanie mające w granicy $\tilde{\theta}_n \rightarrow \theta_n$ następującą postać **obserwowaną równania master**:

$$\frac{\partial p_n(\mathbf{y}_n)}{\partial t_n} + \sum_{i=1}^3 \frac{\partial p_n(\mathbf{y}_n)}{\partial \mathbf{y}_n^i} v_n^i = 0, \quad \text{gdzie} \quad v_n^i = \lim_{\tilde{\theta}_n \rightarrow \theta_n} \hat{v}_n^i \equiv \frac{\tilde{\theta}_n^i - \theta_n^i}{\tilde{\theta}_n^0 - \theta_n^0}, \quad n = 1, 2, \dots, N, \quad (3.69)$$

będącego typem **równania ciągłości strumienia**, gdzie $t_n \equiv \mathbf{y}_n^0$. W (3.69) θ_n^i oraz θ_n^0 są odpowiednio wartościami oczekiwanymi położenia oraz czasu układu.

3.5 Podsumowanie rozważań

Podstawowym przesłaniem wyniesionym z metody estymacyjnej Friedena-Soffera jest to, że TFI jest poprzednikiem Lagrangianu układu [8]. Temat ten rozwinieemy w kolejnym rozdziale. Pewnym minusem teorii Friedena-Soffera mogła wydawać się konieczność zapostulowania nowych zasad informacyjnych. Co prawda z punktu widzenia fenomenologii skuteczność tych zasad w wyprowadzeniu dużej liczby modeli użytecznych do opisu zjawisk wydaje się być całkiem satysfakcjonująca, jednak wyprowadzenie tych zasad przesunęłoby teorię do obszaru bardziej podstawowego. Pozwoliłoby to zarówno na podanie jej przyszłych ograniczeń fenomenologicznych jak i jej możliwych teoretycznych uogólnień.

W tym kontekście, tą właśnie rolę spełnia wyprowadzenie strukturalnej zasady informacyjnej jako konsekwencji analityczności logarytmu funkcji wiarygodności w otoczeniu prawdziwej wartości parametru Θ oraz wskazanie geometrycznego znaczenia informacyjnej zasady wariacyjnej, która leży u podstaw zasady ekstremizacji działania fizycznego. W obecnym rozdziale zwrócono uwagę, że u podstaw informacyjnego zrozumienia zasady wariacyjnej może leżeć idea idealnego pomiaru [10], przy której wariacja pojemności informacyjnej I jest równa (z wyjątkiem znaku) wariacji informacji strukturalnej Q .

W powyższym rozdziale wyprowadzono też równanie master (3.68) dla funkcji wiarygodności, które prowadzi do równania ciągłości strumienia dla punktowego rozkładu w próbie (3.69). Ciekawe jest to, że równanie master pojawia się z rozwinięcia funkcji wiarygodności w szereg Taylora wokół prawdziwej wartości parametru Θ . Siłą rzeczy (por. (3.66)) nie pojawia się więc w nich (logarytmiczna)

⁷Co jest słuszne nawet na poziomie gęstości o ile tylko $P(\tilde{\Theta}) \in \mathcal{S}$ nie posiada w Θ wyższych dżetów niż drugiego rzędu.

część nieliniowa struktury układu. Ta gałąź uogólnienia MNW w klasycznej statystycznej estymacji leży bliżej teorii procesów stochastycznych [34] niż EFI.

Wyprowadzając w Rozdziale 3.1.1 strukturalną zasadę informacyjną [10] wykazano, że metoda Friedena-Soffera jest pewną modyfikacją MNW, pozwalającą, jak się okaże, na nieparametryczną estymację równań ruchu teorii pola lub równań generujących rozkład fizyki statystycznej [8]. Wiele z tych równań otrzymano już w [8] zgodnie z informatycznym zrozumieniem Friedena-Soffera wspomnianym na końcu Rozdziału 3.2. W [16] wyprowadzenia te zostały sprawdzone dla przyjętej w obecnym skrypcie fizycznej postaci zasad informacyjnych [9].

Jednakże dopiero wyprowadzenie strukturalnej zasady informacyjnej pozwala na faktoryzację z obserwowanej SI części, będącej miarą probabilistyczną i w związku z tym na prawidłowe umieszczenie rozkładów spełniających równania różniczkowe metody EFI w odpowiednich podprzestrzeniach przestrzeni statystycznej. Dlatego omówieniu bądź przeliczeniu niektórych rozwiązań EFI z uwzględnieniem tego faktu poświęcimy dwa następne rozdziały.

Należy jednak podkreślić, że Frieden, Soffer i ich współpracownicy Plastino i Plastino, podali metodę rozwiązania układu (różniczkowych) zasad informacyjnych dla problemu EFI, która jest bardzo skuteczna, gdyż poza warunkami brzegowymi i ewentualnymi równaniami ciągłości nie jest ograniczona przez żadną konkretną postać rozkładu. Metodę tą wykorzystamy w dalszym ciągu analizy.

Rozdział 4

Kryteria informacyjne w teorii pola

Główne estymacyjne przesłanie metody EFI. Jak stwierdziliśmy poprzednio, ponieważ podstawowa myśl stojąca za metodą EFI jest następująca: Skoro IF jest infinitesimalnym typem entropii względnej Kulback-Leibler'a, która służy do statystycznego wyboru pomiędzy zaproponowanymi ręcznie modelami, zatem po dodatkowym *ręcznym*, aczkolwiek uzasadnionym, nałożeniu różniczkowych zasad strukturalnych na układ, staje się ona metodą estymującą równania ruchu i ich wyboru drogą wymogu spełnienia zasad¹ informacyjnych. Bądź, jeśli ktoś woli, metoda EFI jest metodą estymującą rozkłady, które są rozwiązaniami tych równań. Jest więc to metoda estymacji nieparametrycznej. Wspomniane równania to np. równania ruchu teorii pola bądź równania generujące rozkłady fizyki statystycznej.

4.1 Informacja Fishera i klasyfikacja modeli

Obecny rozdział poświęcony jest głównie przedstawieniu wstępnej klasyfikacji modeli fizycznych ze względu na skończoność (bądź nieskończoność) pojemności informacyjnej I . Ponadto, poniższe rozważania dla modeli ze skończonym I dotyczą wyłącznie modeli metody EFI. Kolejna, bardziej szczegółowa klasyfikacja pozwala sklasyfikować modele ze względu na wielkość próby N .

Jak pokażemy mechanika klasyczna posiada nieskończoną pojemność informacyjną I . Ściśle mówiąc, mechanika klasyczna jest teorią z symplektyczną strukturą rozmaitości i nie posiada struktury statystycznej. Czasami jednak słyszy się stwierdzenie, że jest ona stochastyczną granicą mechaniki kwantowej. Ale i na odwrót, według von Neumann'a [35] teoria kwantowa jest niespójna z istnieniem zespołów nie posiadających rozmycia (rozproszenia). W związku z tym, dość powszechnie uważa się, że występowanie odstępstw od klasycznego zachowania się układów można uchwycić jedynie na poziomie statystycznym [36].

Poniżej udowodnimy twierdzenie klasycznej statystyki mówiące o niemożliwości wyprowadzenia mechaniki falowej² metody EFI z mechaniki klasycznej. W tym celu wykorzystamy statystyczne pojęcie pojemności informacyjnej, które jest narzędziem dla dwóch sprzężonych z sobą zagadnień, a mianowicie powyżej wspomnianego statystycznego dowodu o niewyprowadzalności mechaniki kwantowej

¹W tym punkcie, działa EFI w stosunku do estymowanego równania ruchu jak MNW w stosunku do estymowanego parametru dla rozkładu znanego typu.

²Wspólnotę mechaniki falowej z kwantową ograniczymy do typów równań różniczkowych zagadnienia Sturm'a-Liouville'a oraz zasady nieoznaczoności Heisenberga. Dowód przeprowadzony w ramach mechaniki falowej w obszarze wspólnym dla obu teorii falowych, uznajemy co najwyżej jako przesłankę jego słuszności w mechanice kwantowej.

z klasycznej i związanego z nim problemu konsystencji samospójnego formalizmu. Ostatni fakt wykorzystywany jest w takich gałęziach badań fizycznych jak nadprzewodnictwo [37], fizyka atomowa i cząstek elementarnych [38] oraz astrofizyka [39].

4.1.1 Podział modeli ze względu na N oraz kategorie I

Jak dotąd nie odnieśliśmy się do wartości N wymiaru próby. Pierwsza klasyfikacja związana z N jest ogólna. Tzn. pokażemy, że modele należą do dwóch różnych, ogólnych kategorii z różną wartością N . Pierwsza z nich posiada skończoną wartość N i jest związana ze skończoną wartością I . Obejmuje ona modele mechaniki falowej i klasycznych teorii pola, gdyż jedno skończenie wymiarowe, polowe rozwiązanie równań ruchu określa ewolucję układu wraz z pełnym określeniem jego struktury w przestrzeni i czasie. Natomiast mechanika klasyczna należy do drugiej kategorii z nieskończonym N , gdyż rozwiązanie równania ruchu nie określa struktury cząstki, która musi być niezależnie od tego równania określona poprzez zdefiniowanie, w każdym punkcie toru cząstki, jej punktowej struktury (np. poprzez dystrybucję δ -Diraca). Mechanika klasyczna okazuje się posiadać nieskończoną wartość I .

4.1.1.1 Dowód podziału na dwie kategorie I

Zobrazujmy powyższe słowa następującą analizą. Dla uproszczenia rozważmy układ jednowymiarowy w położeniu³. Załóżmy wprawdzie, że układ jest opisany przez nieosobliwą dystrybucję. Wtedy dla $N \rightarrow \infty$ pojemność informacyjna I , (2.186), rozbiega się do nieskończoności. Taka sama sytuacja zachodzi jednak dla każdej *osobliwej* dystrybucji jak np. dystrybucja δ -Diraca. Sprawdźmy, że tak jest istotnie. Rozważmy punktową cząstkę swobodną, dla uproszczenia w spoczynku, w położeniu θ , oraz δ -Diracowski ciąg funkcji, np. ciąg funkcji Gaussa:

$$\left\{ \delta_k(y_n|\theta) = \frac{k}{\sqrt{\pi}} \exp(-k^2(y_n - \theta)^2) \right\}, \quad \text{gdzie } k = 1, 2, 3, \dots \quad (4.1)$$

Wtedy, ponieważ dla określonego indeksu k ciągu (4.1), pojemność informacyjna (2.186):

$$I_k = \sum_{n=1}^N \int dy_n \delta_k(y_n|\theta) \left(\frac{\partial \ln \delta_k(y_n|\theta)}{\partial \theta} \right)^2, \quad \text{gdzie } k = 1, 2, 3, \dots \quad (4.2)$$

jest równa:

$$I_k = \frac{N}{\sigma_k^2}, \quad \text{dla } k = 1, 2, 3, \dots, \quad (4.3)$$

gdzie $\sigma_k^2 = \frac{1}{2k^2}$ opisuje wariancję położenia cząstki dla k -tego elementu ciągu (4.1), więc widzimy, że I_k rozbiega się do nieskończoności dla $N \rightarrow \infty$ i nawet jeszcze mocniej, gdy dodatkowo $k \rightarrow \infty$.

Posumowując, dla $N \rightarrow \infty$ pojemność informacyjna I nie istnieje, obojętnie z jaką dystrybucją mielibyśmy do czynienia.

Istnieją więc *dwie, powyżej wymienione, rozdzielne kategorie modeli odnoszące się do wymiaru N próby*. Tzn. dla jednych, takich jak mechanika falowa i teorie pola, N oraz I są skończone, podczas gdy mechanika klasyczna tworzy osobną klasę z nieskończonym N oraz I . To kończy dowód [9] o niewyprowadzalności modeli falowych i teorio-polowych z mechaniki klasycznej.

³Wielowymiarowość czasoprzestrzenna może, w kontekście obecnych rozważań, zmienić co najwyżej znak pojemności informacyjnej.

Powyższy dowód nie obejmuje możliwości wyprowadzenia mechaniki falowej (czy też teorii kwantowych) z klasycznej teorii pola bądź samospójnej teorii pola [38, 40, 41].

Uwaga: Oznacza to, że mechanika klasyczna nie ma skończonego statystycznego pochodzenia⁴, chyba, że tak jak w (4.1) wprowadzi się nieskończoną liczbę statystycznych parametrów, co jednak pociąga za sobą nieskończoność pojemności informacyjnej I .

4.1.2 Podział modeli ze skończonym I na podklasy z różnym N

Jak już wspomnieliśmy, Frieden i Soffer [8] wyprowadzili modele falowe posługując się pojęciem pojemności informacyjnej I oraz zasadami informacyjnymi estymacyjnej metody EFI. Rozwiniemy ten temat w dalszej części obecnego rozdziału. Na razie zauważmy, że stosując zasadniczo⁵ jednocześnie obie zasady informacyjne, strukturalną $\tilde{i} + \tilde{C} + \kappa q = 0$, (3.43), oraz wariacyjną $\delta(I + Q) = 0$, (3.44), oraz uwzględniając odpowiednie fizyczne więzy (wyrażone narzuceniem na układ np. równania ciągłości, symetrii oraz warunków brzegowych), otrzymujemy zróżnicowanie ze względu na N modeli posiadających skończone wartości N oraz I . I tak równanie Kleina-Gordona oraz równanie Schrödingera jako jego nierelatywistyczna granica (por. Dodatek 7.2) posiadają rangę pola $N = 2$, równanie Diraca posiada $N = 8$, równania Maxwell'a posiadają $N = 4$, a teoria grawitacji, zasadniczo bardziej w ujęciu Logunova [42] niż ogólnej teorii względności, posiada $N = 10$ (Dodatek 7.3.2).

4.1.3 Konkluzje i konsekwencje podziału modeli na kategorie I

Powyżej otrzymaliśmy rezultat mówiący, że wszystkie modele opisane strukturalną zasadą informacyjną należą do kategorii skończonej wartości pojemności informacyjnej I oraz, że mechanika klasyczna należy do kategorii nieskończonego I . Zatem w ramach zagadnień rozważanych w skrypcie, granica nie leży pomiędzy tym co micro a makro, ale przebiega pomiędzy teoriami, które mają pochodzenie statystyczne oraz tymi, które mają pochodzenie klasyczno-mechaniczne. Albo lepiej, pomiędzy tym co ma pochodzenie falowe lub szerzej, teorio-polowe, oraz tym co ma pochodzenie ściśle punktowe.

Ponieważ w konstrukcji modeli klasycznej teorii pola oraz mechaniki falowej, użyty jest ten sam statystyczny formalizm informacji Fishera, dlatego jest ona również właściwym narzędziem w konstrukcji samospójnych teorii pola [38], łącząc modele mechaniki falowej i klasycznej teorii pola w jeden, logicznie spójny aparat matematyczny.

Jak wiemy, aby otrzymać jakąkolwiek teorię pola, metoda EFI używa dwóch nowych zasad, wariacyjnej (3.44), która minimalizuje całkowitą fizyczną informację układu oraz obserwowanej (3.43) i oczekiwanej (3.45) zasady strukturalnej, która tą informację zeruje. Frieden i Soffer [8] zwrócili uwagę, że pojęcie informacji poprzedza pojęcie fizycznego działania, a wprowadzony formalizm można słusznie nazwać podejściem Friedena do równań ruchu. Sporo też na tej drodze konstrukcji modeli fizycznych już zrobiono. Jednakże liczne zagadnienia, ze względu na odmienne niż w [8] zrozumienie zasady strukturalnej (patrz [9, 16, 10] oraz obecny skrypt), wymagają ponownego zinterpretowania i zrozumienia. Ciągłe na ogólne opracowanie czeka wprowadzenie do formalizmu informacyjnych poprzedników źródeł oraz lepsze zrozumienie fizyki leżącej u podstaw znaczenia wymiaru próby N .

⁴Frieden co prawda wyprowadził również mechanikę klasyczną z mechaniki falowej, ale jedynie jako graniczny przypadek $\hbar \rightarrow 0$, a wartość N w tym wyprowadzeniu jest nieistotna [8].

⁵Wyjątkiem jest równanie Kleina-Gordona, w wyprowadzeniu którego odwołujemy się jedynie do zasady wariacyjnej, natomiast struktura układu jest narzucona z góry (por. (4.34).)

Poniższe rozważania służą usystematyzowaniu istniejącego już statystycznego aparatu pojęciowego informacji kinetycznej i strukturalnej metody EFI oraz lepszemu opisowi związku informacji fizycznej z całą działaniem.

4.2 Równania różniczkowe metody EFI

Kolejna część obecnego rozdziału poświęcona jest omówieniu rozwiązań zasad informacyjnych metodą EFI dla modeli mechaniki falowej i teorii pola [43]. Punktem wyjścia jest pojemność informacyjna I w jej kinematycznych postaciach (3.51):

$$I = 4 \sum_{n=1}^N \int_{\mathcal{X}} d^4 \mathbf{x} \sum_{\nu=0}^3 \frac{\partial q_n(\mathbf{x})}{\partial \mathbf{x}_\nu} \frac{\partial q_n(\mathbf{x})}{\partial \mathbf{x}^\nu}$$

bądź (3.60):

$$I = 4N \sum_{n=1}^{N/2} \int_{\mathcal{X}} d^4 \mathbf{x} \sum_{\nu=0}^3 \frac{\partial \psi_n^*(\mathbf{x})}{\partial \mathbf{x}_\nu} \frac{\partial \psi_n(\mathbf{x})}{\partial \mathbf{x}^\nu},$$

wyprowadzonych w Rozdziale 3.3, gdzie \mathbf{x}_n^ν są zgodnie z (3.47) przesunięciami wartości pomiarowych położenia zebranych przez układ od ich wartości oczekiwanych. Wyprowadzenie (3.51) oraz (3.60) zostało zaprezentowane w Rozdziale 3.3 i nie zakłada ono (w przeciwieństwie do oryginalnego wyprowadzenia Friedena-Soffera) konieczności istnienia niezmienniczości przesunięcia rozkładów prawdopodobieństwa. Pewne informacje na temat niezmienniczości Lorentzowskiej pojemności informacyjnej I zostały podane w Rozdziale 2.7.1.

Uogólnienia powyższych kinematycznych postaci na przypadek występowania w układzie pól cechowania omówimy w dalszej części rozdziału.

4.2.1 Ogólna postać funkcji gęstości TFI oraz obserwowane zasady informacyjne

Wyprowadzenie strukturalnej zasady informacyjnej zostało przedstawione w Rozdziale 3.1.1. Odwołuje się ono do pełnych danych pomiarowych $(\mathbf{y}_n)_{n=1}^N$, ale jego postać dla przesunięć $(\mathbf{x}_n)_{n=1}^N$ jest dokładnie taka sama [10]. Tak więc, poniżej stosowane zasady informacyjne, strukturalna oraz wariacyjna, będą odwoływały się do miary probabilistycznej $dx p_n(\mathbf{x})$ określonej na przestrzeni przesunięć $\mathbf{x} \in \mathcal{X}$ jako przestrzeni bazowej, gdzie \mathcal{X} jest czasoprzestrzenią Minkowskiego R^4 .

Przystąpmy do przedstawienia konstrukcji mechaniki falowej i teorii pola zgodnie z metodą EFI. Według równania (3.31) TPI została określona jako $K = Q + I$. Ponieważ przesunięcie \mathbf{x}_n nie zależy od parametru θ_m dla $m \neq n$ oraz zakres całkowania dla wszystkich \mathbf{x}_n jest taki sam, dlatego I redukuje się do diagonalnych postaci (3.51) bądź (3.60), a Q do postaci:

$$Q = \sum_{n=1}^N \int_{\mathcal{X}} d^4 \mathbf{x} q_n^2(\mathbf{x}) \mathfrak{F}_n(q_n(\mathbf{x})), \quad (4.4)$$

zgodnie z oznaczeniem w (3.25), bądź w przypadku pola $\psi(\mathbf{x})$, (3.55), do ogólnej (jak zwykle rzeczywistej) postaci:

$$Q \equiv Q_\psi = \int_{\mathcal{X}} d^4 \mathbf{x} \sum_{n,n'=1}^{N/2} \psi_n^*(\mathbf{x}) \psi_{n'}(\mathbf{x}) \mathfrak{F}_{nn'}^\psi(\psi(\mathbf{x}), \psi^*(\mathbf{x}), \psi^{(l)}(\mathbf{x}), \psi^{*(l)}(\mathbf{x})), \quad (4.5)$$

przy czym cała funkcja podcałkowa jest wielomianem pól $\psi(\mathbf{x})$ oraz $\psi^*(\mathbf{x})$, stopnia nie mniejszego niż 2, oraz ich pochodnych rzędu $l = 1, 2, \dots$ (por. (3.19)), natomiast $\mathfrak{F}_{nn'}^\psi$ jest pewną obserwowaną (w ogólności zespoloną) informacją strukturalną układu. Konkretną, jak się okazuje prostą postać Q dla przypadku pól skalarnych Kleina-Gordona oraz pola Diraca omówimy poniżej.

Gęstość TFI: Korzystając z (3.51), (4.4) oraz (3.31), możemy zapisać TFI w postaci:

$$\mathbb{S} \equiv K = \int_{\mathcal{X}} d^4\mathbf{x} k, \quad (4.6)$$

gdzie dla pola opisanego amplitudami q_n :

$$k = 4 \sum_{n=1}^N \left[\sum_{\nu=0}^3 \frac{\partial q_n(\mathbf{x})}{\partial \mathbf{x}_\nu} \frac{\partial q_n(\mathbf{x})}{\partial \mathbf{x}^\nu} + \frac{1}{4} q_n^2(\mathbf{x}) \mathfrak{F}_n(q_n(\mathbf{x})) \right]. \quad (4.7)$$

natomiast dla pola opisanego amplitudami ψ_n :

$$\begin{aligned} k = & 4N \sum_{n,n'=1}^{N/2} \left[\sum_{\nu=0}^3 \delta_{nn'} \frac{\partial \psi_n^*(\mathbf{x})}{\partial \mathbf{x}_\nu} \frac{\partial \psi_{n'}(\mathbf{x})}{\partial \mathbf{x}^\nu} \right. \\ & \left. + \frac{1}{4} \psi_n^*(\mathbf{x}) \psi_{n'}(\mathbf{x}) \mathfrak{F}_{nn'}^\psi(\psi(\mathbf{x}), \psi^*(\mathbf{x}), \psi^{(l)}(\mathbf{x}), \psi^{*(l)}(\mathbf{x})) \right]. \end{aligned} \quad (4.8)$$

Równoważność w (4.6) sugeruje, że K pełni funkcję statystycznego poprzednika (całki) działania \mathbb{S} , natomiast k , będące funkcją gęstości TFI, jest statystycznym poprzednikiem gęstości Lagrangianu \mathcal{L} . Sprawie tej poświęcimy jeden z poniższych rozdziałów.

Uwaga o sformułowaniu Lagrange'a i rząd dżetów funkcji wiarygodności: W dalszej części skryptu założymy, że obserwowana informacja strukturalna nie zawiera pochodnych pól rzędu wyższego niż $l = 1$. Założenie to ma charakter fizyczny. Oznacza ono, że jeśli współrzędne uogólnione (u nas amplitudy) oraz prędkości uogólnione (u nas pochodne amplitud) układu są zadane w pewnej chwili czasu, to ewolucja układu jest całkowicie określona, o ile równania ruchu są 2-giego rzędu. Odpowiada to sformułowaniu Lagrange'a wykorzystywanemu w badaniu dynamicznych i termodynamicznych własności układów.

Fakt ten z punktu widzenia statystycznego oznacza, że interesują nas tylko takie (pod)przestrzenie statystyczne, dla których wszystkie możliwe logarytmy funkcji wiarygodności posiadają r -jety $J_p^r(\mathcal{S}, \mathbf{R})$ w otoczeniu U_p punktu $p \equiv P(\Theta) \in \mathcal{S}$ rzędu $r \leq 2$ (por. Rozdział 2.3). Jest to istotne z punktu widzenia obserwowanej IF (3.4) zdefiniowanej pierwotnie poprzez drugie pochodne logarytmu funkcji wiarygodności po parametrach, co z kolei uwalnia konstrukcję strukturalnej zasady informacyjnej (Rozdział 3.1.1), która jest równaniem metody EFL.

Jednocześnie oczekiwana IF (3.13) wchodzi w nierówność Rao-Cramera, której pewną postacią jest, po dokonaniu w informacji Fishera transformacji Fouriera do przestrzeni pędowej, zasada nieoznaczoności Heisenberga (Dodatek 7.1). Zatem fakt występowania w funkcji Lagrange'a kwadratu pierwszych pochodnych byłby (z tego punktu widzenia) artefaktem konieczności wykonania przez układ estymacji jego czasoprzestrzennych położenia, która to estymacja posiada dolne ogranicze Rao-Cramera na dokładność jednoczesnej estymacji położenia oraz prędkości.

Warunek stałości metryki Rao-Fishera: Odpowiedzmy jeszcze na pytanie, co z punktu widzenia *statystycznego* oznacza niewystępowanie w rozwinięciu $\ln P(\tilde{\Theta})$ w szereg Taylora (3.15) wyrazów rzędu wyższego niż drugi (tzn. brak jetów o rzędzie $r > 2$). Sytuacja ta ma miejsce, gdy obserwowana informacja Fishera iF , (3.4), nie zależy od parametru $\Theta \in V_\Theta$, gdzie V_Θ jest przestrzenią parametru Θ . Wtedy bowiem jej pochodne po parametrze $\Theta = (\theta_n)$ są w równe zero dla każdego punktu $p' = P(\Theta')$ w otoczeniu U_p . Zatem:

$$\frac{\partial}{\partial \Theta'} iF|_{p' \in U_p} = \frac{\partial}{\partial \Theta'} \left(-\frac{\partial^2 \ln P(\Theta')}{\partial \Theta'^2} \right) |_{p' \in U_p} = 0 \Rightarrow j_{p'}^r(\mathcal{S}, R) - j_{p'}^2(\mathcal{S}, R) = 0 \text{ dla } r > 2, \quad (4.9)$$

gdzie $j_p^r(\mathcal{S}, R)$ jest dowolnym elementem przestrzeni jetów $J_p^r(\mathcal{S}, R)$. Lewa strona powyższej implikacji oznacza, że:

$$(iF)_{nn'}|_{p' \in U_p} = \text{const. na } U_p, \quad (4.10)$$

tzn. obserwowana informacja Fishera iF nie zależy w U_p od parametru Θ , z czego wynika *niezależność oczekiwanej IF, czyli metryki Rao-Fishera, od parametru Θ w U_p* .

Takie zachowanie się metryki Rao-Fishera ma następującą ciekawą konsekwencję. Otóż w Rozdziale 5.2.3 okaże się, że fakt *stałości metryki Rao-Fishera* (5.216) jest odpowiedzialny za otrzymanie w ramach metody EFI znanych formuł mechaniki kwantowej (5.232), opisujących splątanie w problemie EPR-Bohm'a.

4.2.1.1 Postać obserwowana zasad informacyjnych

Nałożmy na układ informacyjną obserwowaną zasadę strukturalną $\tilde{i}' + \tilde{C} + \kappa q = 0$, (3.43), oraz informacyjną zasadę wariacyjną $\delta(I + Q) = 0$, (3.44).

Przypadek z amplitudą q : Ze zmodyfikowanej obserwowanej zasady strukturalnej (3.43), biorąc pod uwagę wcześniejsze przejścia pomiędzy (2.31), (2.231), (2.235) oraz (3.51), wynika warunek zerowy:

$$\sum_{\nu=0}^3 \frac{\partial q_n(\mathbf{x})}{\partial \mathbf{x}_\nu} \frac{\partial q_n(\mathbf{x})}{\partial \mathbf{x}^\nu} + \frac{\kappa}{4} q_n^2(\mathbf{x}) \mathbf{qF}_n(q_n(\mathbf{x})) = 0, \quad n = 1, 2, \dots, N. \quad (4.11)$$

Natomiast z zasady wariacyjnej (3.44) wynika układ równań Eulera-Lagrange'a:

$$\sum_{\nu=0}^3 \frac{\partial}{\partial \mathbf{x}^\nu} \left(\frac{\partial k}{\partial (\frac{\partial q_n(\mathbf{x})}{\partial \mathbf{x}^\nu})} \right) = \frac{\partial k}{\partial q_n(\mathbf{x})}, \quad n = 1, 2, \dots, N, \quad (4.12)$$

gdzie k zostało podane w (4.7).

Przypadek z amplitudą ψ : Odpowiednia dla pola ψ obserwowana, zmodyfikowana postać zasady strukturalnej (3.43), która bierze pod uwagę (3.60), jest następująca:

$$\begin{aligned} & \sum_{\nu=0}^3 \frac{\partial \psi_n^*(\mathbf{x})}{\partial \mathbf{x}_\nu} \frac{\partial \psi_n(\mathbf{x})}{\partial \mathbf{x}^\nu} \\ & + \frac{\kappa}{4} \sum_{n'=1}^{N/2} \psi_n^*(\mathbf{x}) \psi_{n'}(\mathbf{x}) \mathbf{qF}_{nn'}^\psi(\psi(\mathbf{x}), \psi(\mathbf{x}), \frac{\partial \psi(\mathbf{x})}{\partial \mathbf{x}}, \frac{\partial \psi^*(\mathbf{x})}{\partial \mathbf{x}}) = 0, \quad n = 1, 2, \dots, N/2. \end{aligned} \quad (4.13)$$

Natomiast układ równań Eulera-Lagrang’a wynikający z zasady wariacyjnej (3.44) jest następujący:

$$\sum_{\nu=0}^3 \frac{\partial}{\partial \mathbf{x}^\nu} \left(\frac{\partial k}{\partial \left(\frac{\partial \psi_n^*(\mathbf{x})}{\partial \mathbf{x}^\nu} \right)} \right) = \frac{\partial k}{\partial \psi_n^*(\mathbf{x})}, \quad n = 1, 2, \dots, N/2, \quad (4.14)$$

gdzie k zostało podane w (4.8).

Zauważmy, że powyższa postać k , (4.8), jest na tyle ogólna, że aby zobaczyć działanie metody EFI wynikające z równań (4.13) i (4.14), należy podać konkretną postać k , dla każdego zagadnienia z polem typu ψ z osobna.

Natomiast, jak się przekonamy, równania (4.11) oraz (4.12) z amplitudami q_n są już zapisane w postaci bliskiej ich bezpośredniego użycia i otrzymania jawnej postaci Φ_n oraz rozwiązań metody EFI, czyli odpowiednich fizycznych równań ruchu (bądź równań generujących, por. Rozdział 5) dla amplitud $q_n(\mathbf{x})$.

Let us summarize this Section. Postacie kinetyczne (3.51) oraz (3.60), oparte o informację Fishera (2.5), są wykorzystywane do konstrukcji równań ruchu (lub równań generujących rozkład) modeli fizycznych. Występują one w (4.11)-(4.14). Natomiast pierwotna postać pojemności I ma swój początek w (3.20) oraz w (3.4) i (3.18). Zgodnie z (3.45), postacie te są równoważne na poziomie oczekiwanym.

Przy samospójnym rozwiązywaniu strukturalnej i wariacyjnej zasady informacyjnej metody EFI, wykorzystywana jest postać obserwowana zasady strukturalnej (3.43), wynikająca z żądania analityczności logarytmu funkcji wiarygodności oraz metryczności przestrzeni statystycznej \mathcal{S} (por. Rozdział 3.1.1). Natomiast oczekiwana strukturalna zasada informacyjna (3.12) jest narzędziem pomocniczym w definicji całkowitej fizycznej informacji K , (3.31), oraz informacyjnej zasady wariacyjnej (3.44).

Poniżej przekonamy się, że wszystkie modelowe różnice leżą po stronie Φ_n , której postać zależy od konkretnego fizycznego scenariusza, włączając w to symetrie oraz warunki brzegowe.

Po raz pierwszy równania (4.11) oraz (4.12) otrzymali Frieden i Soffer [8]. Jednak powyższa ich forma uwzględnia inną interpretację Q (jako obecnej stałe podczas ewolucji układu) oraz faktoryzację probabilistycznego czynnika $q_n^2(\mathbf{x})$ zawartego obligatoryjnie w mierze całkowitej⁶.

4.3 TFI oraz specyficzne formy Q w teorii pola

Poniżej zebrano i rozwinięto wyniki EFI otrzymane poprzednio w [8]. Jednakże zapisano je, szczególnie dla Q , w otwartej formie z punktu widzenia analizy porównawczej modeli [43]. Metoda EFI prowadzi do takiego sformułowania metody teorii pola, która jest zgodna z dzisiejszym opisem mechaniki falowej dla szerokiej klasy struktur.

Jednakże, czasami metoda EFI wraz z całym towarzyszącym jej statystycznym aparatem pojęciowym może doprowadzić do korekty istniejącej analizy. Z sytuacją taką możemy mieć do czynienia np. w przypadku sformułowania zasady nieoznaczoności Heisenberga dla pola *światlnego*. Otóż w świetle nowych eksperymentów, w wyniku których otrzymano *za wąski impuls świetlny w częstotliwości* [44], standardowa Fourierowska podstawa zasady nieoznaczoności jest ostatnio kwestionowana. Wy tłumaczenie istoty nierówności Heisenberga w oparciu o nierówność Rao-Cramera wraz z rozróżnieniem pomiędzy estymacją parametru w przypadku skalarnym [8] (por. Dodatek 7.1) i wektorowym,

⁶Również wyprowadzenie zasady strukturalnej (4.11) czyni ją mniej fundamentalną niż w sformułowaniu Friedena-Soffera [8], a tym co staje się fundamentalne jest funkcja wiarygodności $P(\Theta)$ próbki, czyli łączna gęstość prawdopodobieństwa jej (niewidocznej dla badacza) realizacji.

dla którego zachodzi ciąg nierówności (2.138), może okazać się kluczem do zrozumienia pytań, narastających na skutek nowych eksperymentów ze światłem.

W metodzie EFL, informacja strukturalna Q musi zostać wyprowadzona z użyciem zasady strukturalnej, wariacyjnej i czasami pewnych dodatkowych warunków symetrii, biorących pod uwagę specyficzny fizyczny scenariusz teorii. Szeroki opis metod stosowanych przy rozwiązywaniu równań (4.12) oraz (4.11) można znaleźć w [8].

Jednak poniższe rozważania powinny okazać się pomocne w zrozumieniu metody, szczególnie dla układu zasad informacyjnych (4.13) oraz (4.14) dla pola ψ . Poniżej zostanie pokazane jak możliwe rozwiązania zasad informacyjnych, strukturalnej i wariacyjnej, przewiduje pojawienia się trzech typów pól: N -skalarów [43], fermionów oraz bozonów [8].

4.3.1 Informacja Fouriera

Rozważmy cząstkę jako układ opisany polem rangi N poprzez zbiór zespolonych funkcji falowych $\psi_n(\mathbf{x})$, $n = 1, 2, \dots, N/2$, określonych w czasoprzestrzeni \mathcal{X} położeń $\mathbf{x} \equiv (\mathbf{x}^\mu)_{\mu=0}^3 = (ct, \mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3)$, zgodnie z konstrukcją (3.53) oraz (3.54) z Rozdziału 3.3. Ich transformaty Fouriera $\phi_n(\mathbf{p})$ w sprzężonej do przestrzeni przesunięć \mathcal{X} , energetyczno-pędowej przestrzeni \mathcal{P} czteropędów $\mathbf{p} \equiv (\wp^\mu)_{\mu=0}^3 = (\frac{E}{c}, \wp^1, \wp^2, \wp^3)$, mają postać:

$$\phi_n(\mathbf{p}) = \frac{1}{(2\pi\hbar)^2} \int_{\mathcal{X}} d^4\mathbf{x} \psi_n(\mathbf{x}) e^{i(\sum_{\nu=0}^3 \mathbf{x}^\nu \wp_\nu)/\hbar}, \quad (4.15)$$

gdzie $\sum_{\nu=0}^3 \mathbf{x}^\nu \wp_\nu = Et - \sum_{l=1}^3 \mathbf{x}^l \wp^l$, a \hbar jest stałą Plancka.

Uwaga: Transformacja Fouriera jest *unitarną transformacją zachowującą miarę* na przestrzeni L^2 funkcji całkowalnych z kwadratem, tzn.:

$$\int_{\mathcal{X}} d^4\mathbf{x} \psi_n^*(\mathbf{x}) \psi_m(\mathbf{x}) = \int_{\mathcal{P}} d^4\mathbf{p} \phi_n^*(\mathbf{p}) \phi_m(\mathbf{p}), \quad (4.16)$$

zatem wykorzystując warunek normalizacji prawdopodobieństwa (3.62) otrzymujemy⁷:

$$\sum_{n=1}^{N/2} \int_{\mathcal{X}} d^4\mathbf{x} |\psi_n(\mathbf{x})|^2 = \sum_{n=1}^{N/2} \int_{\mathcal{P}} d^4\mathbf{p} |\phi_n(\mathbf{p})|^2 = 1, \quad (4.18)$$

gdzie $|\psi_n(\mathbf{x})|^2 \equiv \psi_n^*(\mathbf{x}) \psi_n(\mathbf{x})$ oraz $|\phi_n(\mathbf{p})|^2 \equiv \phi_n^*(\mathbf{p}) \phi_n(\mathbf{p})$. Korzystając z (4.15) możemy zapisać I podane wzorem (3.60) w następujący sposób:

$$I[\psi(\mathbf{x})] = I[\phi(\mathbf{p})] = \frac{4N}{\hbar^2} \int_{\mathcal{P}} d^4\mathbf{p} \sum_{n=1}^{N/2} |\phi_n(\mathbf{p})|^2 \left(\frac{E^2}{c^2} - \vec{\wp}^2 \right), \quad (4.19)$$

gdzie $\vec{\wp}^2 = \sum_{k=1}^3 \wp_k \wp^k$.

⁷Równość:

$$\int_{\mathcal{X}} d^4\mathbf{x} \psi_n^*(\mathbf{x}) \psi_n(\mathbf{x}) = \int_{\mathcal{P}} d^4\mathbf{p} \phi_n^*(\mathbf{p}) \phi_n(\mathbf{p}), \quad (4.17)$$

jest treścią twierdzenia Parseval'a.

Określenie kwadratu masy cząstki: Ponieważ I jest z definicji sumą po wartościach oczekiwanych (por. (2.207) i (2.210)), dlatego kwadrat masy cząstki zdefiniowany jako [8]:

$$m^2 := \frac{1}{c^2} \int_{\mathcal{P}} d^4 \mathbf{p} \sum_{n=1}^{N/2} |\phi_n(\mathbf{p})|^2 \left(\frac{E^2}{c^2} - \vec{\wp}^2 \right) \quad (4.20)$$

jest stałą niezależnie od statystycznych fluktuacji energii E oraz pędu $\vec{\wp}$, tzn. przynajmniej wtedy, gdy całkowanie jest wykonane (czyli jako średnia). Tak więc, dla cząstki swobodnej możemy (4.19) zapisać następująco:

$$I[\psi(\mathbf{x})] = I[\phi(\mathbf{p})] = 4N \left(\frac{m c}{\hbar} \right)^2 = const. \quad (4.21)$$

Informacja Fouriera z ψ : Powyższy warunek oznacza, że:

$$K_F \equiv I[\psi(\mathbf{x}^\mu)] - I[\phi(p^\mu)] = 0, \quad (4.22)$$

co korzystając ze stałości $4N \left(\frac{m c}{\hbar} \right)^2$ oraz (4.18) można zapisać jako warunek spełniony przez pole swobodne rangi N :

$$K_F = 4N \int_{\mathcal{X}} d^4 \mathbf{x} \sum_{n=1}^{N/2} \left[\sum_{\nu=0}^3 \left(\frac{\partial \psi_n}{\partial \mathbf{x}_\nu} \right)^* \frac{\partial \psi_n}{\partial \mathbf{x}^\nu} - \left(\frac{m c}{\hbar} \right)^2 \psi_n^* \psi_n \right] = 0. \quad (4.23)$$

Wielkość K_F definiuje tzw. *informację Fouriera* (F), a k_F jej gęstość:

$$k_F = 4N \sum_{n=1}^{N/2} \left[\sum_{\nu=0}^3 \left(\frac{\partial \psi_n}{\partial \mathbf{x}_\nu} \right)^* \frac{\partial \psi_n}{\partial \mathbf{x}^\nu} - \left(\frac{m c}{\hbar} \right)^2 \psi_n^* \psi_n \right]. \quad (4.24)$$

Pomijając fakt, że z powyższych rachunków m^2 wyłania się jako średnia, równanie (4.22), a zatem (4.23), jest odbiciem twierdzenia Parseval's (4.17) i jako takie jest ono zdaniem tautologicznym. Fakt ten oznacza, że transformacja Fouriera odzwierciedla jedynie zmianę bazy w przestrzeni statystycznej \mathcal{S} . Dlatego też, sam z siebie, warunek (4.23) nie nakłada żadnego dodatkowego więzu na układ, *chyba*, że $4N \left(\frac{m c}{\hbar} \right)^2$ jest zadana jako *informacja strukturalna Q układu*. Tylko wtedy (4.23) staje się informacyjną zasadą strukturalną dla układu definiującego szczególny typ pola omawianego w Rozdziale 4.3.2.

4.3.1.1 Informacja Fouriera dla amplitudy rzeczywistej

Rozważmy z kolei cząstkę, jako układ opisany polem rangi N określonym poprzez zbiór rzeczywistych amplitud $q_n(\mathbf{x})$, $n = 1, 2, \dots, N$, na czasoprzestrzeni przesunięć $\mathbf{x} \equiv (\mathbf{x}^\mu)_{\mu=0}^3 \in \mathcal{X}$ i posiadającą pojemność kanału informacyjnego jak w (3.51). Zespolone transformaty Fouriera $\tilde{q}_n(\mathbf{p})$ rzeczywistych funkcji $q_n(\mathbf{x})$, gdzie $\mathbf{p} \equiv (\wp^\mu)_{\mu=0}^3 \in \mathcal{P}$ jest czteropędem, mają postać:

$$\tilde{q}_n(\mathbf{p}) = \frac{1}{(2\pi\hbar)^2} \int_{\mathcal{X}} d^4 \mathbf{x} q_n(\mathbf{x}) e^{i(\sum_{\nu=0}^3 \mathbf{x}^\nu \wp_\nu)/\hbar}. \quad (4.25)$$

Podobnie jak w (4.16), transformacja Fouriera spełnia związek:

$$\int_{\mathcal{X}} d^4 \mathbf{x} q_n^*(\mathbf{x}) q_m(\mathbf{x}) = \int_{\mathcal{P}} d^4 \mathbf{p} \tilde{q}_n^*(\mathbf{p}) \tilde{q}_m(\mathbf{p}). \quad (4.26)$$

Wykorzystując warunek unormowania prawdopodobieństwa:

$$\frac{1}{N} \sum_{n=1}^N \int d^4 \mathbf{x} q_n^2(\mathbf{x}) = 1, \quad (4.27)$$

otrzymujemy, jako konsekwencję twierdzenia Parseval'a:

$$\frac{1}{N} \sum_{n=1}^N \int_{\mathcal{X}} d^4 \mathbf{x} q_n^2(\mathbf{x}) = \frac{1}{N} \sum_{n=1}^N \int_{\mathcal{P}} d^4 \mathbf{p} |\tilde{q}_n(\mathbf{p})|^2 = 1, \quad (4.28)$$

gdzie $|q_n(\mathbf{x})|^2 \equiv q_n^2(\mathbf{x})$ i $|\tilde{q}_n(\mathbf{p})|^2 \equiv \tilde{q}_n^*(\mathbf{p}) \tilde{q}_n(\mathbf{p})$.

Informacja Fouriera dla q : Podobne rachunki jak wykonane poprzednio dla zespolonego pola rangi N , prowadzą w przypadku pola określonego poprzez zbiór N rzeczywistych amplitud $q_n(\mathbf{x})$ i dla pojemności informacyjnej kanału I , (3.51), do następującej jego postaci:

$$I[q(\mathbf{x})] = I[\tilde{q}(\mathbf{p})] = \frac{4}{\hbar^2} \int_{\mathcal{P}} d^4 \mathbf{p} \sum_{n=1}^N |\tilde{q}_n(\mathbf{p})|^2 \left(\frac{E^2}{c^2} - \vec{\wp}^2 \right), \quad (4.29)$$

gdzie $\vec{\wp}^2 = \sum_{k=1}^3 \wp_k \wp_k$.

Określenie kwadratu masy cząstki: Zatem podobnie jak dla pola zespolonego, również dla pola rzeczywistego, kwadrat masy cząstki zdefiniowany następująco:

$$m^2 := \frac{1}{N c^2} \int_{\mathcal{P}} d^4 \mathbf{p} \sum_{n=1}^N |\tilde{q}_n(\mathbf{p})|^2 \left(\frac{E^2}{c^2} - \vec{\wp}^2 \right) \quad (4.30)$$

jest stałą, która nie zależy od statystycznych fluktuacji energii E oraz pędu $\vec{\wp}$. Tak więc i dla cząstki opisanej polem rzeczywistym rangi N , możemy (porównaj (4.21)) zapisać (4.29) w postaci:

$$I[q(\mathbf{x})] = I[\tilde{q}(\mathbf{p})] = 4N \left(\frac{m c}{\hbar} \right)^2 = const., \quad (4.31)$$

co oznacza, że warunek (2.209), $I \geq 0$, pociąga za sobą w zgodzie z (4.31) warunek $m^2 \geq 0$, mówiący o nieobecności tachionów w teorii [30], a wynikający z jej przyczynowości, zgodnie z (2.226) oraz (2.225). Zauważmy, że w zgodzie z (4.30), zerowanie się masy cząstki byłoby niemożliwe dla czasoprzestrzeni z metryką Euklidesową (2.215).

Informacja Fouriera dla pola typu q : Związek (4.31) pozwala na zapisanie informacji Fouriera, w przypadku rzeczywistego swobodnego pola rangi N , w następującej postaci:

$$K_F = 4 \int_{\mathcal{X}} d^4 \mathbf{x} \sum_{n=1}^N \left[\sum_{\nu=0}^3 \frac{\partial q_n}{\partial \mathbf{x}^\nu} \frac{\partial q_n}{\partial \mathbf{x}^\nu} - \left(\frac{m c}{\hbar} \right)^2 q_n^2 \right] = 0. \quad (4.32)$$

Wnioski płynące z zastosowania powyższej postaci informacji Fouriera w przypadku pola bezmasowego, można znaleźć w Dodatku 7.3.1.

4.3.2 Skalary Kleina-Gordona

Rozważmy pole skalarne, którego TFI oznaczmy jako K_S . Równanie ruchu Kleina-Gordona dla swobodnego pola skalarnego rangi N wynika z wariacyjnej zasady informacyjnej (3.44):

$$\delta_{(\psi^*)} K_S = 0, \quad (4.33)$$

gdzie na $k = k_S$, określone ogólnie związkiem (4.8), nałożony jest dodatkowy warunek, wynikający z następującej postaci informacji strukturalnej Q :

$$Q[\phi(\mathbf{p})] = Q_S = \int_{\mathcal{X}} d^4\mathbf{x} q_S \equiv -4N \left(\frac{mc}{\hbar}\right)^2. \quad (4.34)$$

Dla swobodnego skalarnego pola Kleina-Gordona, TFI jest równa jego informacji Fouriera (4.23), tzn. $K = K_S = K_F$. Powyżej q_S jest gęstością informacji strukturalnej dla pola skalarnego:

$$q_S = -4N \left(\frac{mc}{\hbar}\right)^2 \sum_{n=1}^{N/2} \psi_n^* \psi_n. \quad (4.35)$$

Zatem z wariacyjnej zasady informacyjnej $\delta_{(\psi^*)} K_S = 0$ wynika $N/2$ równań Eulera-Lagrange'a (4.14), które dla gęstości TFI równej $k = k_S = k_F$, (4.24), prowadzą do $N/2$ równań Kleina-Gordona⁸. Ich wyprowadzenie można znaleźć w [8, 16].

Uzasadnienie faktu, że pole (ψ_n) z gęstością informacji strukturalnej q_S zadaną przez (4.35) jest polem skalarным, wymaga rozważań związanych z badaniem reprezentacji transformacji izometrii pojemności kanału informacyjnego I , co odkładamy do Rozdziału 4.3.3.2.

Niezmienność Fouriera zasady strukturalnej: Ponieważ dla (4.34) warunek (4.23) stanowi oczekiwaną strukturalną zasadę informacyjną $I + Q = 0$, (3.12), zatem dla swobodnego pola skalarnego transformacja Fouriera jest transformacją unitarną, ze względu na którą warunek (4.23) pozostaje niezmienny.

Masa układu a Fourierowskie splątanie: Powyższy fakt oznacza, że transformacja Fouriera tworzy rodzaj samosplątania pomiędzy reprezentacją położeniową a pędową realizowanych wartości zmiennych układu występujących w I [8], a wyprowadzenie oczekiwanej zasady strukturalnej $I + Q = 0$, (por. (3.12)), jako konsekwencji analityczności logarytmu funkcji wiarygodności [10], wyjaśnia je jako splątanie pędowych stopni swobody układu spowodowane jego masą (4.20).

Uwaga: Podkreślimy, że w przypadku swobodnego pola skalarnego, oczekiwana zasada strukturalna, $I + Q = 0$, jest jedynie odbiciem warunku (4.34). Jest on warunkiem brzegowym i nie jest on rozwiązywany samo-spoźnie wraz z wariacyjną zasadą informacyjną (4.33).

Typy pól skalnych: Podajmy wynikające z powyższej analizy metody EFI dwa typy pól skalnych:

Zwykłe (naładowane) pole skalarne: Pole skalarne mające rangę $N = 2$ ma tylko jedną składową zespoloną, tzn. $\psi \equiv (\psi_n) = \psi_1$ [8]. W przypadku tym informacja strukturalna (4.34) jest równa $Q_S = -8\left(\frac{mc}{\hbar}\right)^2$. Naładowane pole skalarne Higgsa H^+ [45] mogłoby teoretycznie być jego przykładem.

⁸Patrz równanie (4.56) w przypisie.

N-skalary: W przypadku $n > 1$ składowe ψ_n podlegają ewolucji opisanej przez $n = N/2$ nie sprzężonych równań Kleina-Gordona z dwoma dodatkowymi więzami. Pierwszy z nich oznacza, że wszystkie pola ψ_n mają taką samą masę m , a drugim jest warunek normalizacji (4.18). Informacja strukturalna Q takiego układu jest określona przez ogólną postać (4.34) dla pola skalarnego rangi N . Owe skalarne pola Kleina-Gordona rangi N nazwijmy *N-skalarami* [43]. Są one teoretycznie realizowane w ramach tzw. σ -modeli teorii pola [46, 43].

W kolejnym rozdziale omówimy postać TFI oraz Q dla równania Diraca. Podstawowe fakty metody EFI dla pól cechowania w elektrodynamice Maxwella [8] omówione są w Dodatku 7.3.1. Również w Dodatku 7.3.2 zamieszczona jest postać Q w teorii grawitacji [8]. W opracowaniu jest postać Q dla pól nieabelowych [43].

4.3.3 TFI równania Kleina-Gordona dla pól rangi N

Rozdział ten poświęcony jest konstrukcji równania Diraca metodą EFI, z uwzględnieniem pól cechowania⁹. Szczególną uwagę zwrócono na problem kwadratury TFI pola Kleina-Gordona [8, 43].

4.3.3.1 Wstępna foliacja \mathcal{S} oraz pochodna kowariantna. Ogólny zarys problemu

Wybór przestrzeni bazowej członu kinetycznego: Wyjściowa struktura modelu EFI opierała się o analizę wartości oczekiwanej informacji fizycznej $K = I + Q$ na przestrzeni bazowej próby $\mathcal{B} = \mathcal{Y}_1 \times \mathcal{Y}_1 \times \dots \times \mathcal{Y}_N$. Niezależne zmienne losowe Y_n , $n = 1, 2, \dots, N$, przyjmowały wartości $y_n \in \mathcal{Y}_n$. Po przejściu od pełnych danych $y \equiv (y_n)_{n=1}^N \in \mathcal{B}$ do niezależnych zmiennych losowych przesunięć $X_n = Y_n - \theta_n$, $n = 1, 2, \dots, N$, oraz utożsamieniu zbiorów wartości tych zmiennych losowych, tzn. przyjęciu, że $\mathcal{X}_{n=1} \equiv \mathcal{X}$, $n = 1, 2, \dots, N$, zostały skonstruowane kinematyczne postacie pojemności informacyjnej (3.51) oraz (3.60). *Zatem przestrzeni bazową członów kinetycznych jest zbiór przesunięć \mathcal{X} .*

Estymacja na włóknach: Zatem model EFI, z którego wyłoni się model teorii pola jest budowany na *przestrzeni bazowej \mathcal{X}* , będącej w rozważanych przez nas przypadkach czasoprzestrzenią Minkowskiego $\mathcal{X} \equiv R^4$. Pojawiła się ona jako konsekwencja transformacji modelu statystycznego \mathcal{S} z określonego w przestrzeni parametru $V_\Theta \equiv R^4$ do przestrzeni przesunięć \mathcal{X} . Jednakże jednocześnie, tak przed jak i po jego przedefiniowaniu, model statystyczny pozostaje zdefiniowany ponad przestrzenią bazową \mathcal{B} , która jest oryginalną przestrzenią próby. Następnie, w celu uczynienia kinematycznej postaci pojemności informacyjnej I niezmienniczą ze względu na lokalne transformacje cechowania, musimy, poprzez zdefiniowanie pochodnej kowariantnej na przestrzeni bazowej \mathcal{X} , zapisać I w postaci współzmienniczej. Z kolei, zdefiniowanie tej pochodnej kowariantnej oznacza konieczność podania układu współrzędnych na przestrzeni statystycznej \mathcal{S} , co wynika z tego, że pola cechowania są (jak się okaże) amplitudami typu q_n (por. Dodatek 7.3.1).

Zatem wprowadzenie układu współrzędnych na \mathcal{S} nie jest zadaniem trywialnym [6]. Sytuacja ta wynika z konieczności wykonania analizy EFI dla pól rangi N , które z góry przekształcają się zgodnie z transformacjami, których parametry mogą zależeć lokalnie od położenia w przestrzeni bazowej \mathcal{X} . Za-

⁹Zgodnie z uwagą uczynioną powyżej, skrót rezultatów metody EFI dla samych pól cechowania w elektrodynamice Maxwella znajduje się w Dodatku 7.3.1.

tem na \mathcal{X} określana jest *strukturalna grupa symetrii* G wspomnianej powyżej transformacji, która w teorii pola jest grupą Liego pola cechowania. Konstruuje się więc główną wiązkę włóknistą $E(\mathcal{X}, G)$, a fizyczną estymację przeprowadza się na włóknach¹⁰.

Zauważmy, że skoro kwadrat pola cechowania jest elementem \mathcal{S} , więc dla określonej algebry grupy cechowania, wybór cechowania dokonuje częściowej foliacji¹¹ przestrzeni \mathcal{S} na pierwsze warstwy. Dopiero w tym momencie, zasady informacyjne umożliwiają dokonanie wyboru kolejnych foliacji przestrzeni \mathcal{S} na warstwy związane z wszystkimi szczególnymi reprezentacjami grupy G [43], których wymiar jest ściśle związany z rangą pól N [8].

Sens powyższego rozważania ujmijmy tak: Cała procedura EFI musi nie tylko od początku wybrać typ amplitudy (tzn. q_n lub ψ_n), określić zasady informacyjne i warunki brzegowe, ale musi być wykonana od początku we właściwym układzie współrzędnych, który uwzględnia istnienie strukturalnej grupy symetrii G oraz związanych z nią pól cechowania. Pole cechowania umożliwia bowiem wybór podprzestrzeni $T_{\mathcal{P}}H$ przestrzeni stycznej $T_{\mathcal{P}}E$ wiązki głównej $E \equiv E(\mathcal{X}, G)$ w każdym jej punkcie \mathcal{P} , tak, że $T_{\mathcal{P}}E = T_{\mathcal{P}}G \otimes T_{\mathcal{P}}H$, przy czym baza na $T_{\mathcal{P}}H$ może być zdefiniowana jako liniowa kombinacja czterowymiarowej bazy:

$$(\partial_{\mu}) \equiv \left(\frac{\partial}{\partial x^{\mu}} \right) \equiv \left(\frac{\partial}{\partial(ct)}, \frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, \frac{\partial}{\partial x^3} \right) \equiv \left(\frac{\partial}{\partial(ct)}, \vec{\nabla} \right) \quad (4.36)$$

oraz generatorów D_G nieskończonych transformacji grupy G .

Przykład: W przypadku pól cechowania A_{μ} grupy $G = U(1)$ pochodna kowariantna ma postać:

$$D_{\mu} \equiv (D_0, D_l) = \partial_{\mu} - i \frac{e}{c\hbar} A_{\mu}, \quad (4.37)$$

gdzie e jest ładunkiem elektronu. Zatem, wprowadzając pochodną kowariantną do zasad informacyjnych i stosując metodę EFI, otrzymujemy równania ruchu, które rozwiązując dają bazę na znalezionych (pod)rozmaitościach przestrzeni statystycznej \mathcal{S} .

Jest wiele sposobów, na które współzmiennicza postać pojemności informacyjnej I może być odczytana. Poniżej skoncentrujemy się na dwóch z nich, jednej dla pól skalarnych rangi N (tzn. N -skalarów) oraz drugiej, dla pól fermionowych rangi N .

¹⁰ **Uwaga o głównej wiązce (włóknistej):** Mówiąc zwięźle, główna wiązka to taka wiązka włóknista (określona w Rozdziale 2.2.4), której włókno jest strukturalną grupą symetrii G .

Podsumujmy jednak całą dotychczasową informację na temat głównej wiązki włóknistej precyzyjnie: Mając rozmaitość \mathcal{X} oraz grupę Liego G , główna wiązka włóknista $E(\mathcal{X}, G)$ jest rozmaitością taką, że:

1. Grupa G działa na E w sposób różniczkowalny i bez punktów stałych.
2. Przestrzeń bazowa $\mathcal{X} = E/G$, tzn. \mathcal{X} jest przestrzenią ilorazową E względem G , oraz istnieje różniczkowalne odwzorowanie (nazywane rzutowaniem) $\pi : E \rightarrow \mathcal{X}$.
3. Dla każdej mapy $\{U_i\}$ w atlasie dla \mathcal{X} , istnieje różniczkowalne i odwracalne odwzorowanie $\phi_j : \pi^{-1}(U_j) \rightarrow U_j \times G$ zadane przez $E \rightarrow (\pi(\mathcal{P}), f(\mathcal{P}))$ w każdym punkcie $\mathcal{P} \in E$, gdzie $f : \pi^{-1}(U_j) \rightarrow G$ spełnia warunek $f(g\mathcal{P}) = g f(\mathcal{P})$, dla każdego $g \in G$.

Obraz π^{-1} , czyli $U_j \times G$, jest nazywany włóknem. Zatem każde włókno niesie z sobą kopię grupy strukturalnej G .

¹¹ Jako, że baza dla pozostałych pól nie jest jeszcze wybrana.

4.3.3.2 Równanie ruchu Diraca dla pola swobodnego rangi N .

Zgodnie z powyższymi rozważaniami, współzmiennicza forma (3.60) pojemności informacyjnej z pochodną kowariantną D_μ ma postać:

$$I = 4 N \int_{\mathcal{X}} d^4 \mathbf{x} \sum_{n=1}^{N/2} \sum_{\mu=0}^3 (D_\mu \psi_n)^* D^\mu \psi_n . \quad (4.38)$$

Zatem jedyna TFI (3.31) dostępna w metodzie EFI dla równania Kleina-Gordona i każdego pola typu ψ rangi N (skalarne czy fermionowe) ma postać:

$$K = K_{KG} = \int_{\mathcal{X}} d^4 \mathbf{x} k_{KG} \equiv 4 N \int_{\mathcal{X}} d^4 \mathbf{x} \sum_{n=1}^{N/2} \sum_{\mu=0}^3 (D_\mu \psi_n)^* D^\mu \psi_n + Q , \quad (4.39)$$

gdzie k_{KG} jest gęstością TFI dla równania Kleina-Gordona (por. (4.8)).

Podobnie, wykorzystując (3.51) w miejsce (3.60) moglibyśmy zapisać K_{KG} dla pola bosonowego, do czego powrócimy jednak później.

Są dwie drogi, którymi analiza oparta o TFI zadaną przez (4.39) może podążać [8, 43]. Pierwsza związana jest z N -skalarami a druga z polami Diraca.

TFI Kleina-Gordona

Wtedy, gdy rozważamy pole N -skalara, to jak wiemy Q jest równe Q_S zadanemu przez (4.34). Fakt ten oznacza, że TFI dla równania ruchu N -skalara, sprowadza się do postaci Kleina-Gordona, tzn.:

$$K = K_S = K_{KG} \quad \text{dla} \quad Q = Q_S , \quad (4.40)$$

a wariacyjna zasada informacyjna:

$$\delta_{(\psi^*)} K_S \equiv \delta_{(\psi^*)} (I + Q_S) = 0 , \quad (4.41)$$

zadana przez (3.41), a w konsekwencji przez (4.14), prowadzi do równania Kleina-Gordona dla pola skalarnego [8, 16] (patrz (4.56)).

TFI Diraca

Jak wiemy, dla pola Diraca, równanie Kleina-Gordona jest otrzymane drogą kwadratury równania Diraca [47]. Zatem, mogłoby się wydawać, że metoda informacyjna nie wybiera sama z siebie właściwej postaci TFI dla pola fermionowego. Sytuacja ma się jednak zgoła inaczej. Przedstawimy ją poniżej [43].

Transformacje izometrii I : Zapiszmy (4.38) pól ψ , zarówno dla N -skalara jak i pola fermionowego rangi N , w następującej postaci:

$$I = 4N \sum_{n=1}^{N/2} \int_{\mathcal{X}} d^4 \mathbf{x} \sum_{\mu, \nu=0}^3 (D_\mu \psi_n^*(\mathbf{x})) \eta^{\mu\nu} (D_\nu \psi_n(\mathbf{x})) , \quad (4.42)$$

która to postać pozwala zauważyć, że jest ona niezmiennicza ze względu na transformacje izometrii działające w $N/2$ - wymiarowej, zespolonej przestrzeni $\mathcal{C}^{N/2}$ pól $\psi \equiv (\psi_n(\mathbf{x}))$ rangi N .

Dwa typy izometrii I :

Przypadek skalarów: W przypadku N -skalarów ψ tworzących podprzestrzeń w $\mathcal{C}^{N/2}$ izometrie te są transformacjami identycznościowymi, pozostawiając pola ψ niezmiennymi.

Przypadek pól Diraca: Natomiast dla pól fermionowych rangi N określonych na przestrzeni bazowej \mathcal{X} Minkowskiego, transformacje izometrii tworzą grupę Clifforda $Pin(1, 3)$, będącą podzbiorem algebry Clifforda $C(1, 3)$. Elementy grupy Clifforda $Pin(1, 3)$ działają w podprzestrzeni $\mathcal{C}^{N/2}$ spinorów ψ .

Macierze Diraca: Okazuje się, że macierze Diraca γ^μ , $\mu = 0, 1, 2, 3$, (por. (4.48)-(4.50)), tworzą spinorową reprezentację ortogonalnej bazy w $C(1, 3)$ i spełniają tożsamości:

$$\eta^{\mu\nu} = 1/2 \{ \gamma^\mu, \gamma^\nu \}, \quad \mu, \nu = 0, 1, 2, 3, \quad (4.43)$$

które są podstawowym związkiem dla iloczynu Clifforda. W przypadku spinorów rangi $N = 8$, baza w $Pin(1, 3)$ jest $2^{N/2} = 16$ wymiarowa [47].

Faktoryzacja K_{KG} : Dla trywialnego przypadku N -skalarów, postać (4.39) jest formą podstawową. Jednak w przypadku spinorów rangi N , po skorzystaniu z (4.43) w (4.42), można dokonać rozkładu fizycznej informacji K_{KG} , (4.39), na składowe. Postać jawnego rozkładu K_{KG} na składowe i ich faktoryzację, gdzie każdy z otrzymanych czynników jest elementem grupy Clifforda $Pin(1, 3)$, podał Frieden [8].

Główny rezultat rozkładu fizycznej informacji K_{KG} , z uwzględnieniem pól cechowania w pochodnej kowariantnej D_μ , który czyni zadość "ograniczeniu" Kleina-Gordona (co oznacza, że wymnożenie czynników i dodanie ich dałoby na powrót (4.39)), ma postać:

$$K = K_D \equiv 4N \int_{\mathcal{X}} d^4\mathbf{x} \sum_{n=1}^{N/2} \sum_{\mu=0}^3 (D_\mu \psi_n)^* D^\mu \psi_n + Q, \quad (4.44)$$

gdzie informacja strukturalna Q jest równa:

$$Q = Q_D = \int_{\mathcal{X}} d^4\mathbf{x} q_D \equiv -4N \int_{\mathcal{X}} d^4\mathbf{x} \sum_{n=1}^{N/2} \left[v_{1n} v_{2n} + \left(\frac{mc}{\hbar} \right)^2 \psi_n^* \psi_n \right] + (\text{pozostałe człony}), \quad (4.45)$$

gdzie q_D jest zgodnie z (3.23) gęstością informacji strukturalnej, a $\frac{N}{2} = 4$ wymiarowe wektory kolumnowe $v_i = (v_{i1}, v_{i2}, v_{i3}, v_{i4})^T$, $i = 1, 2$, mają składowe:

$$v_{1n} = \sum_{n'=1}^4 \left(i \mathbf{1} D_0 - \beta \frac{mc}{\hbar} + \sum_{l=1}^3 i \alpha^l D_l \right)_{nn'} \psi_{n'}, \quad n = 1, 2, 3, 4 \quad (4.46)$$

oraz

$$v_{2n} = \sum_{n'=1}^4 \left(-i \mathbf{1} D_0 + \beta^* \frac{mc}{\hbar} + \sum_{l=1}^3 i \alpha^{l*} D_l \right)_{nn'} \psi_{n'}^*, \quad n = 1, 2, 3, 4, \quad (4.47)$$

gdzie macierze α^l , $l = 1, 2, 3$, oraz β są macierzami Diraca (4.49)¹², a $\mathbf{1}$ jest 4×4 - wymiarową macierzą jednostkową.

Zasady informacyjne dla równania Diraca: W końcu, informacyjna zasada strukturalna (3.43) dla $\kappa = 1$ i gęstości informacji strukturalnej q_D określonej zgodnie z (4.45) oraz zasada wariacyjna (3.44) mają postać:

$$\tilde{i}' + \tilde{\mathbf{C}} + q_D = 0 \quad \text{dla} \quad \kappa = 1 \quad \text{oraz} \quad \delta_{(\psi^*)} K_D = \delta_{(\psi^*)} (I + Q_D) = 0. \quad (4.51)$$

Powyższe zasady dają na poziomie obserwowanym warunki (4.13) oraz (4.14) metody EFI, czyli układ dwóch równań różniczkowych, które są rozwiązywane samospójnie. W wyniku otrzymujemy równanie Diraca:

$$v_1 = \left(i D_0 - \beta \frac{m c}{\hbar} + \sum_{l=1}^3 \alpha^l i D_l \right) \psi = 0, \quad \text{gdzie} \quad \psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix}. \quad (4.52)$$

Generacja masy: Zwróćmy uwagę, że w trakcie powyższego rozkładu K_{KG} z członu kinetycznego I , (4.42), generowane są wszystkie składniki informacji strukturalnej Q , (4.45). W trakcie tej procedury człon masowy generowany jest poprzez sprzężenie Fourierowskie (4.15) pomiędzy reprezentacją położeniową i pędową oraz uśrednienie dokonane w reprezentacji energetyczno-pędowej, tak jak to miało miejsce dla (4.20). Zatem również masa pola Diracowskiego jest przejawem istnienia Fourierowskiego samosplątania pomiędzy reprezentacją położeniową a pędową, a cała procedura jest odbiciem:

- i) założenia analityczności logarytmu funkcji wiarygodności,
- ii) podziału na część Fisherowską I oraz (początkowo nieznaną) część strukturalną Q z wysumowaniem po kanałach informacyjnych i uśrednieniem po przestrzeni próby \mathcal{B} opisanym powyżej (3.12),
- iii) przejścia z pojemnością informacyjną I do Friedenowskiej postaci kinematycznej z całkowaniem po zakresie przesunięć \mathcal{X} amplitud rozkładu układu, opisanym w Rozdziale 3.3,
- iv) tautologicznego wygenerowania, zgodnie z zasadą Macha, informacji strukturalnej Q z informacji Fishera I , z uwzględnieniem niezmienniczości I ze względu na transformację amplitud będącą jej izometrią oraz transformację Fouriera pomiędzy reprezentacją położeniową i pędową.

Przedstawiony schemat generacji masy nie obejmuje wyznaczenia jej wartości. Mówi on raczej o tym czego jej pojawienie się jest wyrazem.

¹²Macierze Diraca γ^μ , $\mu = 0, 1, 2, 3$, występujące w tzw. kowariantnej formie równania Diraca:

$$(i \gamma^\mu D_\mu - m) \psi = 0, \quad (4.48)$$

wyrażają się poprzez macierze Diraca β oraz α^l , $l = 1, 2, 3$, w sposób następujący: $\gamma^0 = \beta$ oraz $\gamma^l = \beta \alpha^l$, ($\gamma_\mu = \sum_{\nu=0}^3 \eta_{\mu\nu} \gamma^\nu$). Macierze Diraca $\vec{\alpha} \equiv (\alpha^1, \alpha^2, \alpha^3)$ oraz β mają postać:

$$\alpha^l = \begin{pmatrix} 0 & \sigma^l \\ \sigma^l & 0 \end{pmatrix}, \quad l = 1, 2, 3, \quad \beta = \begin{pmatrix} \mathbf{1} & 0 \\ 0 & -\mathbf{1} \end{pmatrix} \quad (4.49)$$

gdzie $\sigma^l = \sigma_l$, ($l = 1, 2, 3$) są macierzami Pauliego, a $\mathbf{1}$ jest macierzą jednostkową:

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \mathbf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (4.50)$$

Warunek zerowania się pozostałych członów: W równaniach (4.46) oraz (4.47) macierz jednostkowa 1 , która stoi przy D_0 oraz $\frac{N}{2} \times \frac{N}{2} = 4 \times 4$ wymiarowe macierze Diraca β , α_l , są jednymi z elementów grupy Clifforda $Pin(1, 3)$. Jak wspomnieliśmy, Frieden [8] przeprowadził opisaną dekompozycję K_{KG} , (4.39), do postaci K_D podanej w (4.44). Równocześnie pokazał, że wyrażenie oznaczone w (4.45) jako “*pozostałe człony*” zeruje się przy założeniu, że *macierze β i α_l spełniają relacje algebry Clifforda*.

Podsumowanie: Zauważyliśmy, że pojemność informacyjna I w TFI jest zadana przez (4.42) zarówno dla pól N -skalarów jak i pól fermionowych rangi N . Kluczową sprawą jest, że dla każdego równania ruchu, określona jest tylko jemu charakterystyczna postać q . Zauważyliśmy, że dla N -skalarów, jedynym członem, który tworzy całkowitą fizyczną informację jest K_{KG} w postaci (4.39) bez żadnego ukrytego rozkładu i faktoryzacji, pochodzących z nie-Fourierowskiego splątania. W przypadku równania Diraca, informacja fizyczna Kleina-Gordona K_{KG} , (4.39), również wchodzi w całkowitą informację fizyczną Diraca K_D , ale tylko jako jej szczególna część (jak to można również zauważyć z porównania (4.44), (4.45) oraz (4.39) i (4.34)), splątana Fourierowsko w K_F (4.23). Fakt ten jest blisko związany z efektem EPR-Bohm’a [48] opisanym w Rozdziale 5.2.2.

Postać Q dla równania Diraca: Tak więc, w zgodzie z (4.45), szczegółowa postać Q dla równania ruchu Diraca okazała się być inna niż dla N -skalarów i można ją zapisać następująco:

$$Q = Q_D \equiv \mathbb{S}_q - 4N \int_{\mathcal{X}} d^4\mathbf{x} \sum_{n=1}^{N/2} \left[\left(\frac{mc}{\hbar} \right)^2 \psi_n^* \psi_n \right], \quad (4.53)$$

gdzie \mathbb{S}_q jest charakterystyczną częścią Diracowską działania dla kwadratury równania Diraca, wyznaczoną dla pola Diraca o randze $N = 8$ przy wzięciu pod uwagę wszystkich symetrii układu.

Całka działania kwadratury równania Diraca: Ponieważ jednak całka działania (por. Rozdział 4.3.5) dla kwadratury specyficznej części Diracowskiej zachowuje się następująco:

$$\mathbb{S}_q = -4N \int_{\mathcal{X}} d^4\mathbf{x} \sum_{n=1}^{N/2} (v_{1n} v_{2n}) + (\text{pozostałe człony}) = 0, \quad (4.54)$$

tzn. spełnia warunek zerowy, zatem $K \equiv K_D$ redukuje się i (w kwadraturze) określa postać K_{KG} dla (4.39)¹³:

$$K = K_{KG} \equiv 4N \int_{\mathcal{X}} d^4\mathbf{x} \sum_{n=1}^{N/2} \left[\sum_{\mu=0}^3 (D_\mu \psi_n)^* D^\mu \psi_n - \left(\frac{mc}{\hbar} \right)^2 \psi_n^* \psi_n \right]. \quad (4.57)$$

¹³ Równanie Kleina-Gordona otrzymane z wariacji informacji (4.57) ma postać ($\vec{\nabla} = (\partial/\partial x^l)$, $l = 1, 2, 3$):

$$-c^2 \hbar^2 \left(\vec{\nabla} - \frac{ie\vec{A}}{c\hbar} \right) \cdot \left(\vec{\nabla} - \frac{ie\vec{A}}{c\hbar} \right) \psi_n + \hbar^2 \left(\frac{\partial}{\partial t} + \frac{ie\phi}{\hbar} \right)^2 \psi_n + m^2 c^4 \psi_n = 0. \quad (4.55)$$

Dla pola swobodnego czteropotencjał cechowania $A_\mu = (\phi, -\vec{A})$ jest równy zeru i wtedy otrzymujemy:

$$-c^2 \hbar^2 \nabla^2 \psi_n + \hbar^2 \frac{\partial^2}{\partial t^2} \psi_n + m^2 c^4 \psi_n = 0. \quad (4.56)$$

Uwaga o fundamentalnej różnicy pola skalarnego i pola Diraca: Ważną sprawą jest zauważenie, że chociaż postać I dla N -skalarów oraz pól fermionowych rangi N wygląda “powierzchniowo” tak samo, to jednak układy te różnią się istotnie. W samej bowiem rzeczy, podczas gdy N -skalar jest rozwiązaniem równania ruchu, które wynika jedynie z wariacyjnej zasady informacyjnej (4.41), to pole fermionowe jest samospójnym rozwiązaniem zasad informacyjnych, zarówno wariacyjnej jak i strukturalnej, podanych w (4.51).

4.3.4 Końcowe uwagi o wkładzie Q w zasadę strukturalną

Wynik Rozdziału 3.1.1 [10] związany z wyprowadzeniem strukturalnej zasady informacyjnej jest ogólny, o ile tylko, w celu zagwarantowania słuszności rozwinięcia w szereg Taylora, funkcja log-wiarygodności $\ln P(\Theta)$ jest analityczną funkcją wektorowego parametru położenia Θ w zbiorze jego wartości V_Θ . Z kolei, w Rozdziale 3.2 zauważyliśmy, że model rozwiązany przez metodę EFI jest modelem metrycznym z metryką Rao-Fisher’a oraz, że na poziomie całkowym, model metryczny jest równoważny modelowi analitycznemu.

Dla układów, które nie posiadają dodatkowych więzów różniczkowych, wszystkie położeniowe stopnie swobody są związane jedynie poprzez analizę modelu metrycznego, wynikającą z równań (3.43)-(3.44) oraz wspomnianą zasadę Macha, generującą człon strukturalny z Fisherowskiego członu kinematycznego, co prowadzi do czynnika efektywności $\kappa = 1$. Równania Kleina-Gordona oraz Diraca omawiane w Rozdziałach 4.3.2 oraz 4.3.3 są modelami tego typu.

Jeśli na układ nałożony jest dodatkowy warunek, który nie wynika ani ze strukturalnej zasady informacyjnej (3.43), $\tilde{i}' + \tilde{C} + q = 0$ z $\kappa = 1$, ani z zasady wariacyjnej (3.44), wtedy wzrasta związek strukturalny pomiędzy położeniowymi stopniami swobody co powoduje, że κ wiążąca w równaniu (3.43) gęstość informacji strukturalnej q z gęstością pojemności informacyjnej \tilde{i}' , musi maleć poniżej wartości 1. Własność ta wynika z wklęsłości pojemności informacyjnej I przy mieszaniu układów [8], które pojawiają się np. na skutek wprowadzenia dodatkowego różniczkowego więzu. Na przykład, omawiany w Dodatku 7.3.1 warunek Lorentza dla pola cechowania Maxwella, który jest równaniem typu równania ciągłości strumienia, pojawia się nie jako konsekwencja równań ruchu badanego pola, ale jako ograniczenie szukane na drodze niezależnej statystycznej estymacji. Najprawdopodobniej ten dodatkowy warunek¹⁴ pojawia się jako rezultat rozwinięcia w szereg Taylora “gołej” funkcji wiarygodności $P(\Theta)$ [10], podobnie jak wyprowadzone w Rozdziale 3.4 równanie master (co sygnalizowałoby spójność całej statystycznej metody estymacyjnej¹⁵). Zatem warunek Lorentza nałożony na układ redukuje jego symetrię, co pociąga za sobą pojawienie się zasady strukturalnej $\tilde{i}' + \tilde{C} + \kappa q = 0$, (3.43), z czynnikiem efektywności κ mniejszym niż 1. W przypadku równań Maxwella omówionych w Dodatku 7.3.1, zwrócimy uwagę [43], że wartość $\kappa = 1/2$ pojawia się automatycznie w metodzie EFI na skutek samospójnego rozwiązania równania strukturalnego, wariacyjnego i nałożonego warunku Lorentza.

4.3.5 Zasada najmniejszego działania v.s. zasady informacyjne

Powyższe rozważania Rozdziału 4.3 są częściowo poświęcone omówieniu dodatkowego wyniku zastosowania metody EFI, tzn. ustaleniu różnych postaci informacji strukturalnej Q . Zauważono, że

¹⁴Jednak to przypuszczenie należy udowodnić.

¹⁵Z drugiej strony, same równania master leżą w innej części klasycznej statystycznej estymacji, tzn. w obszarze działania teorii procesów stochastycznych [34].

przy umiarkowanie rozbudowanym aparacie metody EFI, następuje nie tylko wyprowadzenie, tzn. estymacja, jej wynikowego równania ruchu (lub równania generującego rozkład), ale jakby przy okazji, pojawienie się postaci obserwowanej informacji strukturalnej $\mathfrak{F}_n(q_n(\mathbf{x}))$.

Konstrukcja całki działania: Z powyższego faktu wynika określenie związku pomiędzy całką działania wraz z zasadą najmniejszego działania, a całkowitą informacją fizyczną wraz z zasadami informacyjnymi. Okazuje się bowiem, że po wstawieniu $\mathfrak{F}_n(q_n(\mathbf{x}))$ z powrotem do K otrzymujemy całkę działania \mathbb{S} modelu [43]:

$$\mathbb{S}(q_n(\mathbf{x}), \partial q_n(\mathbf{x})) = K(q_n(\mathbf{x}), \partial q_n(\mathbf{x}), \mathfrak{F}_n(q_n(\mathbf{x}))). \quad (4.58)$$

Sens równości (4.58) określa poniższa konstrukcja $\mathbb{S}(q_n(\mathbf{x}), \partial q_n(\mathbf{x}))$. Wpierw rozwiązujemy zasady informacyjne, strukturalną i wariacyjną, znajdując obserwowaną informację strukturalną $\mathfrak{F}_n(q_n(\mathbf{x}))$. Następnie możemy otrzymać równanie ruchu na dwa sposoby:

(a) Pierwszy sposób analizy: W metodzie EFI wstawiamy otrzymane $\mathfrak{F}_n(q_n(\mathbf{x}))$ od razu do równań *Eulera-Lagrange'a* wynikających z wariacyjnej zasady informacyjnej z gęstością informacji fizycznej $k(q_n(\mathbf{x}), \partial q_n(\mathbf{x}), \mathfrak{F}_n(q_n(\mathbf{x})))$.

(b) Drugi sposób analizy: W drodze do całki działania teorii pola wstawiamy otrzymane $\mathfrak{F}_n(q_n(\mathbf{x}))$ do $K(q_n(\mathbf{x}), \partial q_n(\mathbf{x}), \mathfrak{F}_n(q_n(\mathbf{x})))$ i stosując zasadę najmniejszego działania dla tak skonstruowanego $\mathbb{S}(q_n(\mathbf{x}), \partial q_n(\mathbf{x}))$, otrzymujemy równanie *Eulera-Lagrange'a* teorii pola.

Rezultat: Otrzymane w konsekwencji analizy typu (a) lub (b) równanie ruchu, bądź równanie generujące rozkład¹⁶, jest w obu podejściach takie samo¹⁷.

Struktura statystyczna teorii: Jednakże, z powodu definicji \mathbb{S} danej przez lewą stronę równania (4.58), pojawia się pytanie, czy metoda EFI, która wykorzystuje K dane przez prawą stronę równania (4.58), daje jakieś dodatkowe informacje w porównaniu do zasady najmniejszego działania dla \mathbb{S} . Potwierdzająca odpowiedź jest następująca. Według Rozdziału 4.3 metoda EFI dokonuje (zazwyczaj) rozkładu K na podstawowe bloki podając otwarcie wszystkie fizyczne i statystyczne pojęcia, i narzędzia analizy. Równość (4.58), która jest jedynie definicją \mathbb{S} , wyraża bowiem jednocześnie fakt, że K ma bardziej złożoną strukturę niż \mathbb{S} .

Powyższe stwierdzenie oznacza, że \mathbb{S} oraz zasada najmniejszego działania niosą zarówno mniejszą fizyczną informację niż oryginalne zasady informacyjne (uzupełnione fizycznymi więzami modelu),

¹⁶W przypadku wyprowadzenia równania Boltzmanna w Rozdziale 5.1.3, po rozwiązaniu układu równań strukturalnego (5.24) i wariacyjnego (5.17), otrzymuje się postać (5.30) na $\mathfrak{F}_n(q_n(\mathbf{x}))$. Jest to postać, która po wstawieniu do równania wariacyjnego (5.17) daje równanie generujące (5.34) amplitudę $q_n(\mathbf{x})$. Zasada wariacyjna (5.14) (gdzie $\mathfrak{F}_n(q_n(\mathbf{x}))$ wyznaczone w (5.30) występuje w k , (5.15)), dałaby również równanie generujące (5.34).

Równość (4.58) należy więc rozumieć następująco. Znając $\mathbb{S}(q_n(\mathbf{x}), \partial q_n(\mathbf{x}))$, które występuje po lewej stronie równania (4.58) jako K ze wstawioną, samospójnie wyznaczoną postacią $\mathfrak{F}_n(q_n(\mathbf{x}))$ i wariując je dopiero wtedy ze względu na $q_n(\mathbf{x})$, otrzymujemy to samo równanie generujące co rozwiązanie obu zasad strukturalnej i wariacyjnej jednocześnie.

¹⁷Natomiast w ogólności, nie można przyrównać samej zasady wariacyjnej metody EFI dającej równanie Eulera-Lagrange'a (dla amplitudy $q_n(\mathbf{x})$), w którym może być uwikłana obserwowana informacja strukturalna $\mathfrak{F}_n(q_n(\mathbf{x}))$, do zasady najmniejszego działania \mathbb{S} , która daje końcową postać równania Eulera-Lagrange'a dla $q_n(\mathbf{x})$ bez żadnego, bezpośredniego śladu postaci $\mathfrak{F}_n(q_n(\mathbf{x}))$.

jak i niższą informację o pojęciach *statystycznych*.

Zmniejszenie informacji statystycznej: Po otrzymaniu rozwiązania równań informacyjnych i “ściągnięciu” wyniku do \mathbb{S} , znikają z \mathbb{S} nie tylko pojęcia informacji Fishera oraz automatycznie pojemności informacyjnej, lecz i pojęcie całkowitej wewnętrznej dokładności modelu, tzn. informacji Stama.

Zmniejszenie informacji fizycznej: W końcu, jeśli rzeczywistość (w znaczeniu matematycznych podstaw modelowych) leżąca poza modelami teorii pola byłaby statystyczna, wtedy odrzucając ich oryginalny związek z analizą na przestrzeni statystycznej \mathcal{S} oraz estymacyjną metodą EFL, która wybiera fizycznie właściwą (pod)przestrzeń \mathcal{S} , tracimy również oryginalne narzędzia tej analizy służące do konstrukcji działania \mathbb{S} . Zamiast tego zadowalamy się zgadywaniem jego postaci jedynie na podstawie fizycznych warunków wstępnych, zastanawiając się np. jak szczególnym jest pojęcie amplitudy oraz jak wyjątkowa jest zasada nieoznaczoności Heisenberga.

Przykład: Obok uwagi na wstępie Rozdziału 4.3, ostatnie eksperymenty ze światłem wydają się mówić, że Fourierowskie częstości nie są widoczne w optycznej lokalizacji najmniejszych jego impulsów. W takiej sytuacji zasada Heisenberga $\delta\nu\delta t \geq 1/2$ mogłaby nie reprezentować sobą żadnej granicznej fizycznej rzeczywistości, gdzie $\delta\nu$ oraz δt są poprzez transformację Fouriera związanymi z sobą szerokościami połówkowymi impulsu w częstości oraz w czasie. Gdyby eksperymenty te potwierdziły się, wtedy zasada Heisenberga straciłaby swoje oparcie w transformacji Fouriera dla zmiennych komplementarnych [44]. Natomiast jest możliwe, że znalazłaby ona wtedy swoje oparcie w nierówności Rao-Cramera, która podaje dolne ograniczenie na wariancję estymatora parametru w przypadku pomiaru jednokanałowego [8] (por. Dodatek 7.1).

Rozdział 5

Przykłady z fizyki statystycznej i ekonofizyki oraz efekt EPR-Bohm’a

Ogólne statystyczne podstawy estymacji MNW zostały przedstawione w Rozdziałach 1-2. Natomiast w Rozdziałach 3 oraz 4 skryptu przedstawiono metodę EFI jako szczególny typ estymacyjnej procedury statystycznej, opracowanej w ramach teorii pomiaru. Metoda EFI pokazuje w jaki sposób wychodząc z pojęcia funkcji wiarygodności oraz pojemności informacyjnej I (por. Rozdział 3.3) otrzymać więzy strukturalne wynikające z analitycznej informacyjnej zasady strukturalnej, która wraz z wariacyjną zasadą informacyjną oraz równaniem ciągłości (lub ogólniej, równaniem typu master por. Rozdział 2.2.3)), prowadzi do równań różniczkowych teorii [8]. W trakcie procedury otrzymujemy informację strukturalną Q opisywanego układu, która wraz z pojemnością informacyjną I tworzy informację fizyczną układu K , będącą statystycznym poprzednikiem całki działania (por. Rozdział 4.3.5). Wyprowadzenia równań ruchu lub równań generujących rozkład zasadzają się na potraktowaniu wszystkich warunków nałożonych na układ jako związków na odchylenia (fluktuacje) wartości pomiarowych od wartości oczekiwanych. Przy tym, analizowane dane pojawiają się jako efekt pomiaru dokonanego przez układ (por. Rozdział 2.8).

Obecny Rozdział dzieli się na dwie części, pierwszą mającą zastosowanie termodynamiczne oraz drugą, opisującą zjawisko EPR-Bohm’a. Za wyjątkiem krótkiej analizy rozwiązań metody EFI dla równań transportu Boltzmann’a, obie części łączy wymiar próby $N = 1$.

5.1 Wyprowadzenie klasycznej fizyki statystycznej z informacji Fishera

Celem obecnego rozdziału jest wyprowadzenie metodą EFI podstawowych rozkładów klasycznej fizyki statystycznej. Otrzymamy więc równania generujące, z których wyprowadzone zostaną: rozkład Boltzmann’a dla energii, a następnie rozkład Maxwella-Boltzmann’a dla pędu. Jako przykład zastosowania analizy w ekonofizyce, podamy przykład produktywności branż w modelu Aoki-Yoshikawy. Przedstawione rachunki idą śladem analizy Friedena, Soffera, Plastino i Plastino [8], jednak zostaną one wykonane w oparciu o wprowadzoną w Rozdziale 3.1.1 strukturalną zasadę informacyjną [10]. Różnicę interpretacyjną pomiędzy obu podejściami podano w Rozdziale 3.2.

Dodatkowo wyprowadzony zostanie warunek informacyjny na górne ograniczenie tempa wzrostu entropii Shannona [8].

5.1.1 Fizyczne sformułowanie zagadnienia

Rozważmy gaz składający się z M identycznych cząsteczek o masie m zamknięty w zbiorniku. Temperatura gazu ma stałą wartość T . Ruch cząsteczek jest losowy i oddziałują one ze sobą poprzez siły potencjalne, zderzając się ze sobą i ściankami naczynia, przy czym zakładamy, że są to zderzenia sprężyste. Średnia prędkość każdej cząsteczki jest równa zero.

Oznaczmy przez $\theta_\varphi = (\theta_\epsilon, \vec{\theta}_\varphi)$ czterowektor wartości oczekiwanej energii oraz pędu cząsteczki, gdzie indeks φ oznacza pęd. We współrzędnych kartezjańskich $\vec{\theta}_\varphi = (\theta_{\varphi 1}, \theta_{\varphi 2}, \theta_{\varphi 3})$. Podobnie jak w (3.47) wprowadzamy zmienną losową $Y = (Y_\epsilon \equiv \frac{E}{c}, \vec{Y}_\varphi)$, przyjmującą wartości $\mathbf{y} = (\mathbf{y}_\epsilon \equiv \epsilon/c, \vec{\mathbf{y}}_\varphi)$, które składowe spełniają związki:

$$\mathbf{y}_\epsilon \equiv \frac{\epsilon}{c} = \theta_\epsilon + \mathbf{x}_\epsilon, \quad \frac{\epsilon_0}{c} \leq \mathbf{y}_\epsilon \leq \infty \quad (5.1)$$

$$\vec{\mathbf{y}}_\varphi = \vec{\theta}_\varphi + \vec{\mathbf{x}}_\varphi, \quad \vec{\mathbf{y}}_\varphi = (\mathbf{y}_{\varphi 1}, \mathbf{y}_{\varphi 2}, \mathbf{y}_{\varphi 3}), \quad (5.2)$$

gdzie

$$\vec{\mathbf{x}}_\varphi = \vec{\varphi} \quad (5.3)$$

oznacza fluktuację pędu, natomiast c oznacza prędkość światła. Zmienne i parametry energetyczne \mathbf{y}_ϵ , \mathbf{x}_ϵ oraz θ_ϵ zostały wyrażone w jednostce współrzędnych pędowych *energia/c*. Parametry θ_ϵ oraz $\vec{\theta}_\varphi$ są odpowiednimi wartościami oczekiwanymi energii (z dokładnością do $1/c$) oraz pędu, a \mathbf{x}_ϵ oraz $\vec{\mathbf{x}}_\varphi$ fluktuacjami względem wartości oczekiwanych.

Znajdziemy rozkład prawdopodobieństwa dla fluktuacji energii X_ϵ przyjmującej wartości \mathbf{x}_ϵ oraz fluktuacji pędu \vec{X}_φ przyjmującej wartości $\vec{\varphi}$ dla jednej, dowolnej cząsteczki gazu w dowolnej chwili czasu t . Ponieważ wartość t nie musi być duża, zatem rozważamy gaz, który nie koniecznie jest w stanie równowagi. Będziemy więc szukać postaci nierównowagowego rozkładu prawdopodobieństwa, odpowiadającego tak postawionemu problemowi.

Uwaga o czterowektorze energii-pędu: Jednak po pierwsze, współrzędne czterowektora fluktuacji energii-pędu $(X_\epsilon, \vec{X}_\varphi)$ nie są statystycznie niezależne, tzn. nie są niezależnymi stopniami swobody układu. Po drugie, jak się okaże, spełniają one zasadę dyspersyjną typu (4.20), więc nie tworzą układu zmiennych Fishera (porównaj (2.223)). Zatem ogólny problem wymagałby estymacji odpowiednich równań generujących dla skomplikowanego czasoprzestrzennego zagadnienia, tzn. należałoby wyznaczyć łączny rozkład prawdopodobieństwa współrzędnych $(X_\epsilon, \vec{X}_\varphi)$, co wykracza poza zakres skryptu. Niemniej np. w przypadku relatywistycznych zjawisk astrofizycznych, taka estymacja może okazać się niezbędna.

W skrypcie ograniczymy się jedynie do wyznaczenia brzegowych rozkładów prawdopodobieństwa dla X_ϵ (i w konsekwencji dla E) oraz dla \vec{X}_φ , co w nierelatywistycznej granicy jest uzasadnione.

5.1.2 Informacja kinetyczna i strukturalna oraz sformułowanie zasad informacyjnych

Tak więc, statystycznie jedna cząsteczka podlega łącznemu rozkładowi prawdopodobieństwa $p(\mathbf{x}_\epsilon, \vec{\mathbf{x}}_\varphi)$, przy czym współrzędne czterowektora pędu nie są niezależne. Uproszczona analiza skon-

centruje się na rozkładach brzegowych, których analiza ze względu na wspomniany brak niezależności nie odtwarza analizy łącznej, chociaż jest słuszna w przybliżeniu nierelatywistycznym.

Wyznamy więc *brzegowe amplitudy prawdopodobieństwa* $q(\mathbf{x}_\epsilon)$ oraz $q(\vec{\mathbf{x}}_\varphi)$:

$$q_n(\mathbf{x}_\epsilon), \quad n = 1, \dots, N_\epsilon \quad \text{oraz} \quad q_n(\vec{\mathbf{x}}_\varphi), \quad n = 1, \dots, N_\varphi, \quad (5.4)$$

a następnie powrócimy do brzegowych rozkładów prawdopodobieństwa $p(\mathbf{x}_\epsilon)$ oraz $p(\vec{\mathbf{x}}_\varphi)$. W końcu po wyznaczeniu $p(\mathbf{x}_\epsilon)$ skorzystamy z (5.1), aby otrzymać wymagany rozkład $p(\epsilon)$. Podobnie, korzystając z (5.2) otrzymamy po wyznaczeniu $p(\vec{\mathbf{x}}_\varphi)$ rozkład prędkości $p(\vec{\mathbf{y}}_\varphi)$, przy czym te dwa ostatnie są równe w naszych rozważaniach, ze względu na średnią wartość prędkości cząsteczki $\vec{\theta}_\varphi = 0$.

Pojemność informacyjna dla parametrów czterowektor energii-pędu: Chociaż $(\mathbf{x}_\epsilon, \vec{\mathbf{x}}_\varphi)$ nie jest czterowektorem we wspomnianym ujęciu Fishera, to jest on czterowektorem w sensie Lorentza. Korzystamy więc z metryki czasoprzestrzeni Minkowskiego w postaci $(\eta_{\nu\mu}) = \text{diag}(1, -1, -1, -1)$ zgodnie z (2.214).

W ogólnym przypadku rozkładu łącznego oraz próbkowania czterowektora energii i pędu, pojemność informacyjna ma dla parametru wektorowego $\Theta = ((\theta_{\nu n})_{\nu=0}^3)_{n=1}^N$ postać (2.231)¹:

$$I = \sum_{n=1}^N \int_{\mathcal{B}} dy P(y|\Theta) \left[\left(\frac{\partial \ln P(y|\Theta)}{\partial \theta_{\epsilon n}} \right)^2 - \sum_{k=1}^3 \left(\frac{\partial \ln P(y|\Theta)}{\partial \theta_{\varphi k n}} \right)^2 \right], \quad (5.5)$$

gdzie $y = (\mathbf{y})_{n=1}^N$ jest N -wymiarową próbą, a \mathcal{B} przestrzenią próby. Zatem pojemności informacyjne $I(\Theta_\epsilon)$ oraz $I(\Theta_\varphi)$ dla rozkładów brzegowych fluktuacji energii oraz pędu mają postać²:

$$I(\Theta_\epsilon) = 4 \int_{\mathcal{X}_\epsilon} d\mathbf{x}_\epsilon \sum_{n=1}^{N_\epsilon} \left(\frac{dq_n(\mathbf{x}_\epsilon)}{d\mathbf{x}_\epsilon} \right)^2 \quad (5.6)$$

oraz

$$I(\Theta_\varphi) = -4 \int_{\mathcal{X}_\varphi} d\vec{\mathbf{x}}_\varphi \sum_{n=1}^{N_\varphi} \sum_{k=1}^3 \left(\frac{dq_n(\vec{\mathbf{x}}_\varphi)}{dx_{\varphi k}} \right)^2, \quad (5.7)$$

gdzie minus w (5.5) i w konsekwencji w (5.7), wynika zgodnie z Rozdziałem 2.7.1 z uwzględnienia metryki Minkowskiego (2.214), natomiast \mathcal{X}_ϵ oraz \mathcal{X}_φ są zbiorami wartości odpowiednio zmiennych X_ϵ oraz X_φ . Z poniższych rachunków przekonamy się, że nieuwzględnienie metryki Minkowskiego (gdy $\eta_{00} = 1$ to $\eta_{kk} = -1$ dla $k = 1, 2, 3$), doprowadziłoby w konsekwencji w termodynamicznych rozważaniach do błędnego rozkładu prędkości³.

¹W zgodzie z ogólną własnością kontrakcji indeksu Minkowskiego dla czterowektora $x \equiv (x_\nu)_{\nu=0}^3 = (x_0, x_1, x_2, x_3)$:

$$\sum_{\nu\mu=0}^3 \eta_{\nu\mu} \frac{df}{dx_\nu} \frac{df}{dx_\mu} = \left(\frac{df}{dx_0} \right)^2 - \sum_{k=1}^3 \left(\frac{df}{dx_k} \right)^2.$$

W [8] metryka ma postać Euklidesową, a zmienne mają urojoną współrzędną przestrzenną $(x_0, i\vec{x})$, por. Rozdział 2.7.1.

²Jak w Rozdziale 3.3, indeks n przy współrzędnej pominięto, korzystając z założenia, że rozproszenie zmiennej nie zależy od punktu pomiarowego próby.

³Odpowiedź na pytanie, czy rozważania termodynamiczne są przyczyną metryki Minkowskiego niezbędnej w relatywistycznej teorii pola, wykracza poza obszar skryptu. Niemniej autor skryptu uważa, że tak się istotnie sprawy mają, tzn. że przestrzeń Euklidesowa z transformacją Galileusza są pierwotne wobec przestrzeni Minkowskiego z transformacją Lorentza. Stąd podejście efektywnej teorii pola Logunova [42] do teorii grawitacji jest bliższe teorii pomiaru fizycznego Friedena-Soffera (którą jest EFI). Nieco więcej na ten temat można znaleźć w Dodatku 7.3.2.

Zasady informacyjne: Zasady informacyjne, strukturalna (3.43) oraz wariacyjna (3.44) mają poniższą postać. Dla energii:

$$\tilde{i}'(\Theta_\epsilon) + \tilde{\mathbf{C}}_\epsilon + \kappa q(\Theta_\epsilon) = 0, \quad \delta_{(q_n)} (I(\Theta_\epsilon) + Q(\Theta_\epsilon)) = 0, \quad (5.8)$$

oraz dla pędu:

$$\tilde{i}'(\Theta_{\vec{\varphi}}) + \tilde{\mathbf{C}}_{\vec{\varphi}} + \kappa q(\Theta_{\vec{\varphi}}) = 0, \quad \delta_{(q_n)} (I(\Theta_{\vec{\varphi}}) + Q(\Theta_{\vec{\varphi}})) = 0. \quad (5.9)$$

Gęstości pojemności informacyjnych, $\tilde{i}'(\Theta_\epsilon)$ oraz $\tilde{i}'(\Theta_{\vec{\varphi}})$, są określone zgodnie z (3.42) i (3.43), a gęstości informacji strukturalnych, $q(\Theta_\epsilon)$ oraz $q(\Theta_{\vec{\varphi}})$, są określone zgodnie z (3.23), natomiast $Q(\Theta_\epsilon)$ oraz $Q(\Theta_{\vec{\varphi}})$ są odpowiednimi informacjami strukturalnymi.

W pierwszej kolejności rozważymy problemem (5.8) dla $p(\epsilon)$.

Przypomnienie roli zasad informacyjnych: W rachunkach metody EFI prowadzących do równania generującego rozkład, obok zasady wariacyjnej (3.44) wykorzystywana jest postać (3.43) zmodyfikowanej obserwowanej zasady strukturalnej. Warto pamiętać, że zasada (3.43) wynika z żądania istnienia rozwinięcia Taylora logarytmu funkcji wiarygodności wokół prawdziwej wartości parametru (por. Rozdział 3.1.1) oraz z metryczności przestrzeni statystycznej \mathcal{S} . Natomiast oczekiwana strukturalna zasada informacyjna (3.12) jest narzędziem pomocniczym w definicji całkowitej fizycznej informacji K , (3.31), oraz informacyjnej zasady wariacyjnej (3.44).

5.1.3 Rozkład Boltzmanna dla energii

Poniżej podamy rozwiązanie zasad informacyjnych (5.8), strukturalnej oraz wariacyjnej, otrzymując w pierwszym kroku analizy równanie generujące amplitudy dla rozkładu Boltzmanna.

Założmy wstępnie, że wartość fluktuacji energii \mathbf{x}_ϵ zmienia się w pewnym zakresie $\langle \mathbf{x}_\epsilon^{min}, \mathbf{x}_\epsilon^{max} \rangle$:

$$\mathbf{x}_\epsilon^{min} \leq \mathbf{x}_\epsilon \leq \mathbf{x}_\epsilon^{max}. \quad (5.10)$$

W ten sposób wprawdzie uchwycimy ogólną zależność amplitudy rozkładu od $\mathbf{x}_\epsilon^{max}$, a następnie dokonamy przejścia granicznego, przechodząc z górną granicą fluktuacji energii $\mathbf{x}_\epsilon^{max}$ do nieskończoności.

Pojemność informacyjna ma postać (3.51):

$$I(\Theta_\epsilon) = 4 \int_{\mathbf{x}_\epsilon^{min}}^{\mathbf{x}_\epsilon^{max}} d\mathbf{x}_\epsilon \sum_{n=1}^N q_n'^2(\mathbf{x}_\epsilon), \quad \text{gdzie} \quad q_n'(\mathbf{x}_\epsilon) \equiv \frac{dq_n(\mathbf{x}_\epsilon)}{d\mathbf{x}_\epsilon}, \quad (5.11)$$

gdzie rozkłady prawdopodobieństwa p_n dla fluktuacji energii są powiązane z amplitudami q_n zależnością (2.168):

$$p_n(\mathbf{x}_\epsilon) = q_n^2(\mathbf{x}_\epsilon). \quad (5.12)$$

Informacja strukturalna zgodnie z (3.25) jest następująca:

$$Q(\Theta_\epsilon) = \int_{\mathbf{x}_\epsilon^{min}}^{\mathbf{x}_\epsilon^{max}} d\mathbf{x}_\epsilon q(\Theta_\epsilon) = \int_{\mathbf{x}_\epsilon^{min}}^{\mathbf{x}_\epsilon^{max}} d\mathbf{x}_\epsilon \sum_{n=1}^N q_n^2(\mathbf{x}_\epsilon) \Phi_n(q_n(\mathbf{x}_\epsilon)). \quad (5.13)$$

Informacyjna zasada wariacyjna w (5.8) dla pojemności informacyjnej I , (5.11), oraz informacji strukturalnej Q , (5.13), ma postać:

$$\delta_{(q_n)} K = \delta_{(q_n)} (I(\Theta_\epsilon) + Q(\Theta_\epsilon)) = \delta_{(q_n)} \left(\int_{\mathbf{x}_\epsilon^{min}}^{\mathbf{x}_\epsilon^{max}} d\mathbf{x}_\epsilon k \right) = 0, \quad (5.14)$$

gdzie k jest równe:

$$k = 4 \sum_{n=1}^N \left(q_n'^2 + \frac{1}{4} q_n^2(\mathbf{x}_\epsilon) \Phi_n(q_n(\mathbf{x}_\epsilon)) \right), \quad (5.15)$$

zgodnie z ogólną postacią gęstości obserwowanej informacji fizycznej (4.7) dla amplitud q_n .

Rozwiązaniem problemu wariacyjnego (5.14) względem $q_n(\mathbf{x}_\epsilon)$ jest *równanie Eulera-Lagrange'a* (4.12):

$$\frac{d}{d\mathbf{x}_\epsilon} \left(\frac{\partial k}{\partial q_n'(\mathbf{x}_\epsilon)} \right) = \frac{\partial k}{\partial q_n(\mathbf{x}_\epsilon)} \quad \text{dla } n = 1, 2, \dots, N, \quad (5.16)$$

gdzie 'prim' oznacza pochodną po \mathbf{x}_ϵ , tzn. $q_n'(\mathbf{x}_\epsilon) \equiv dq_n(\mathbf{x}_\epsilon)/d\mathbf{x}_\epsilon$.

Zatem dla rozważanego problemu, równanie (5.16) dla k jak w (5.15), przyjmuje postać:

$$2 q_n''(\mathbf{x}_\epsilon) = \frac{\partial(\frac{1}{4} q_n^2(\mathbf{x}_\epsilon) \Phi_n(q_n(\mathbf{x}_\epsilon)))}{\partial q_n} = \frac{d(\frac{1}{4} q_n^2(\mathbf{x}_\epsilon) \Phi_n(q_n(\mathbf{x}_\epsilon)))}{dq_n}. \quad (5.17)$$

Zmodyfikowana obserwowana zasada strukturalna: Po wycalkowaniu (5.11) przez części, pojemność I wynosi:

$$I(\Theta_\epsilon) = -4 \int_{\mathbf{x}_\epsilon^{min}}^{\mathbf{x}_\epsilon^{max}} d\mathbf{x}_\epsilon \sum_{n=1}^N q_n(\mathbf{x}_\epsilon) q_n''(\mathbf{x}_\epsilon) + C, \quad (5.18)$$

gdzie

$$C \equiv 4 \int_{\mathbf{x}_\epsilon^{min}}^{\mathbf{x}_\epsilon^{max}} d\mathbf{x}_\epsilon \left(q_n(\mathbf{x}_\epsilon) q_n'(\mathbf{x}_\epsilon) \right)' = 4 \sum_{n=1}^N C_n \quad (5.19)$$

oraz

$$C_n = q_n(\mathbf{x}_\epsilon^{max}) q_n'(\mathbf{x}_\epsilon^{max}) - q_n(\mathbf{x}_\epsilon^{min}) q_n'(\mathbf{x}_\epsilon^{min}). \quad (5.20)$$

Zatem widzimy, że *zmodyfikowana obserwowana zasada strukturalna* w (5.8) jest ze względu na (5.18)-(5.19) oraz (5.13), następująca:

$$\sum_{n=1}^N \left(-q_n(\mathbf{x}_\epsilon) q_n''(\mathbf{x}_\epsilon) + \tilde{C}_n + \kappa \frac{1}{4} q_n^2(\mathbf{x}_\epsilon) \Phi_n(q_n(\mathbf{x}_\epsilon)) \right) = 0, \quad (5.21)$$

gdzie

$$\tilde{C}_n = C_n / (\mathbf{x}_\epsilon^{max} - \mathbf{x}_\epsilon^{min}). \quad (5.22)$$

Zatem na poziomie obserwowanym, dla każdego $n = 1, 2, \dots, N$, otrzymujemy zasadę strukturalną w postaci⁴:

$$-q_n(\mathbf{x}_\epsilon)q_n''(\mathbf{x}_\epsilon) + \tilde{C}_n + \kappa \frac{1}{4}q_n^2(\mathbf{x}_\epsilon) \mathfrak{F}_n(q_n(\mathbf{x}_\epsilon)) = 0. \quad (5.24)$$

Korzystając z (5.17) w (5.24) otrzymujemy:

$$\frac{1}{2}q_n \frac{d(q_n^2(\mathbf{x}_\epsilon) \mathfrak{F}_n(q_n(\mathbf{x}_\epsilon)))}{dq_n} = \kappa q_n^2(\mathbf{x}_\epsilon) \mathfrak{F}_n(q_n(\mathbf{x}_\epsilon)) + 4\tilde{C}_n. \quad (5.25)$$

Poniżej, dla uproszczenia zapisu pominiemy przy rozwiązywaniu równania (5.25) oznaczenie argumentu \mathbf{x}_ϵ w amplitudzie q_n . Zapiszmy (5.25) w postaci:

$$2 \frac{dq_n}{q_n} = \frac{d \left[\frac{1}{4}q_n^2 \mathfrak{F}_n(q_n) \right]}{\kappa \left[\frac{1}{4}q_n^2 \mathfrak{F}_n(q_n) \right] + \tilde{C}_n}. \quad (5.26)$$

Rozwiązując powyższe równanie różniczkowe otrzymujemy kolejno:

$$2 \ln q_n + \alpha_n' = \frac{1}{\kappa} \ln \left(\frac{\kappa}{4} q_n^2 \mathfrak{F}_n(q_n) + \tilde{C}_n \right)$$

i po przekształceniu:

$$2\kappa \ln \alpha_n'' q_n = \ln \left(\frac{\kappa}{4} q_n^2 \mathfrak{F}_n(q_n) + \tilde{C}_n \right),$$

gdzie α_n' jest w ogólności zespoloną stałą całkowania oraz stała $\alpha_n'' = \exp(\alpha_n'/2)$, skąd:

$$q_n^2(\mathbf{x}_\epsilon) \mathfrak{F}_n(q_n) = \frac{4}{\kappa} \left(\alpha_n^2 q_n^{2\kappa} - \tilde{C}_n \right), \quad (5.27)$$

gdzie (w ogólności zespolona) stała $\alpha_n^2 = \alpha_n''^{2\kappa}$. Zatem w rezultacie otrzymaliśmy obserwowaną informację strukturalną, $\mathfrak{F}_n(q_n(\mathbf{x}_\epsilon))$, jako znaną funkcję amplitudy $q_n(\mathbf{x}_\epsilon)$.

Równanie generujące z κ : W końcu, korzystając z (5.17) oraz z (5.27), otrzymujemy równanie różniczkowe dla $q_n(\mathbf{x}_\epsilon)$:

$$q_n''(\mathbf{x}_\epsilon) = \alpha_n^2 q_n^{2\kappa-1}(\mathbf{x}_\epsilon). \quad (5.28)$$

Podsumowanie: Wariacyjna oraz strukturalna obserwowana zasada informacyjna zostały zapisane w postaci układu równań (5.17) oraz (5.24), który rozwiązując, dał w rezultacie szukaną postać (5.28) równania generującego amplitudę $q_n(\mathbf{x}_\epsilon)$. Znalezienie postaci tego równania było pośrednim celem metody EFL.

Warunek analityczności i metryczności: Z postaci (5.28) widać, że skoro amplituda $q_n(\mathbf{x}_\epsilon)$ jest funkcją rzeczywistą, zatem stała α_n^2 musi być rzeczywista. Rozwiązanie równania (5.28) znajdziemy wtedy, gdy współczynnik efektywności wynosi:

$$\kappa = 1, \quad (5.29)$$

⁴Gdyby nie wyciąkać w (5.18) I przez części, wtedy z (3.40) i po skorzystaniu z postaci kinematycznej pojemności, zasada strukturalna na poziomie obserwowanym miałaby postać:

$$q_n'^2(\mathbf{x}_\epsilon) + \frac{\kappa}{4}q_n^2(\mathbf{x}_\epsilon) \mathfrak{F}_n(q_n(\mathbf{x}_\epsilon)) = 0. \quad (5.23)$$

czyli dla przypadku, gdy zasada strukturalna dla układu jest konsekwencją analityczności logarytmu funkcji wiarygodności (3.1) oraz metryczności przestrzeni statystycznej \mathcal{S} .

W przypadku $\kappa = 1$ obserwowana informacja strukturalna (5.27) przyjmuje prostą postać:

$$q_n^2(\mathbf{x}_\epsilon) \Phi_n(q_n(\mathbf{x}_\epsilon)) = 4(\alpha_n^2 q_n^2(\mathbf{x}_\epsilon) - \tilde{C}_n) , \quad (5.30)$$

a całkowa postać SI jest następująca:

$$Q(\Theta_\epsilon) = 4 \int d\mathbf{x}_\epsilon \sum_{n=1}^N \alpha_n^2 q_n^2(\mathbf{x}_\epsilon) - 4 \sum_{n=1}^N C_n , \quad (5.31)$$

co po skorzystaniu z unormowania kwadratu amplitudy:

$$\int d\mathbf{x}_\epsilon q_n^2(\mathbf{x}_\epsilon) = \int d\mathbf{x}_\epsilon p_n(\mathbf{x}_\epsilon) = 1 , \quad (5.32)$$

daje:

$$Q(\Theta_\epsilon) = 4 \sum_{n=1}^N (\alpha_n^2 - C_n) = 4 \sum_{n=1}^N \alpha_n^2 - C . \quad (5.33)$$

W ostatnim przejściu w (5.33) skorzystano z postaci stałej $C \equiv 4 \sum_{n=1}^N C_n$ wprowadzonej w (5.18).

Równanie generujące: Dla rozważanego przypadku $\kappa = 1$, równanie generujące (5.28) ma postać:

$$q_n''(\mathbf{x}_\epsilon) = \alpha_n^2 q_n(\mathbf{x}_\epsilon) . \quad (5.34)$$

Sprawdzenie rachunków: Wstawiając (5.34) do (5.18), otrzymujemy następującą postać pojemności informacyjnej:

$$I(\Theta_\epsilon) = -4 \sum_{n=1}^N \alpha_n^2 + C , \quad (5.35)$$

co wraz z (5.33) oznacza sprawdzenie poprawności rachunku, poprzez spełnienie przez otrzymane rozwiązanie oczekiwanego strukturalnego warunku $I + Q = 0$, zgodnie z (3.45).

Rozwiązanie równania generującego: Kolejnym etapem analizy jest rozwiązanie równania generującego (5.34).

Najogólniejsza postać amplitudy $q_n(\mathbf{x}_\epsilon)$ będącej rozwiązaniem równania (5.34) jest przy warunku $\mathbf{x}_\epsilon^{min} \leq \mathbf{x}_\epsilon \leq \mathbf{x}_\epsilon^{max}$, jak w (5.10), następująca:

$$q_n(\mathbf{x}_\epsilon) = B_n \exp(\alpha_n \mathbf{x}_\epsilon) + D_n \exp(-\alpha_n \mathbf{x}_\epsilon) , \quad \mathbf{x}_\epsilon^{min} \leq x \leq \mathbf{x}_\epsilon^{max} , \quad B_n, D_n = const. \quad (5.36)$$

Ponieważ stała α_n^2 jest rzeczywista, zatem wprowadzając nową rzeczywistą stałą β_n , można α_n przedstawić jako $\alpha_n = \beta_n$ i wtedy rozwiązanie (5.36) ma charakter czysto eksponencjalny:

$$q_n(\mathbf{x}_\epsilon) = B_n \exp(\beta_n \mathbf{x}_\epsilon) + D_n \exp(-\beta_n \mathbf{x}_\epsilon) , \quad (5.37)$$

bądź jako $\alpha_n = i\beta_n$, i wtedy rozwiązanie (5.36) ma charakter czysto trygonometryczny:

$$q_n(\mathbf{x}_\epsilon) = B_n \exp(i\beta_n \mathbf{x}_\epsilon) + D_n \exp(-i\beta_n \mathbf{x}_\epsilon) . \quad (5.38)$$

Warunek normalizacji dla amplitud: Tak określone funkcje muszą być dopuszczalne jako amplitudy prawdopodobieństwa, zatem muszą spełniać *warunek normalizacji* dla gęstości prawdopodobieństwa (5.32).

Założmy, że wartość fluktuacji energii \mathbf{x}_ϵ nie jest ograniczona od góry, co zrealizujemy jako dążenie $\mathbf{x}_\epsilon^{max}$ do nieskończoności. Jednak kwadrat funkcji trygonometrycznej nie może być unormowany do jedności dla $\mathbf{x}_\epsilon^{max} \rightarrow \infty$, zatem funkcja trygonometryczna (5.38) nie jest dopuszczalnym rozwiązaniem.

Pozostaje więc rozwiązanie eksponencjalne (5.37). Ponieważ jednak warunek unormowania (5.32) ma być zadany na przedziale otwartym $\mathbf{x}_\epsilon^{min} \leq \mathbf{x}_\epsilon < \infty$, zatem część rozwiązania z dodatnią eksponentą musi być odrzucona ze względu na jej rozbieżność do nieskończoności. Skąd otrzymujemy żądanie, że dla $\beta_n \geq 0$ stała $B_n = 0$.

Podsumowując, szukana postać amplitudy jest więc następująca:

$$q_n(\mathbf{x}_\epsilon) = D_n \exp(-\beta_n \mathbf{x}_\epsilon), \quad \beta_n \in \mathbf{R}_+, \quad \mathbf{x}_\epsilon^{min} \leq \mathbf{x}_\epsilon < \infty. \quad (5.39)$$

Z powyższego i z warunku normalizacji (5.32) $\int_{\mathbf{x}_\epsilon^{min}}^{\infty} d\mathbf{x}_\epsilon q_n^2(\mathbf{x}_\epsilon) = 1$, wyznaczamy stałą D_n , otrzymując:

$$D_n = \sqrt{2\beta_n} \exp(\beta_n \mathbf{x}_\epsilon^{min}). \quad (5.40)$$

Ostateczna postać amplitudy: Rozwiązanie to zostało otrzymane dla przypadku $\beta_n = \alpha_n \in \mathbf{R}_+$, zatem ostateczną postacią (5.36) jest:

$$q_n(\mathbf{x}_\epsilon) = \sqrt{2\alpha_n} \exp[\alpha_n(\mathbf{x}_\epsilon^{min} - \mathbf{x}_\epsilon)], \quad \alpha_n \in \mathbf{R}_+. \quad (5.41)$$

Zauważmy, że α_n jest w jednostkach $[c/energia]$.

Końcowa postać pojemności informacyjnej: W końcu możemy wyznaczyć pojemność informacyjną. Wstawiając (5.41) do (5.11), otrzymujemy:

$$I(\Theta_\epsilon) = 4 \sum_{n=1}^N \alpha_n^2 > 0, \quad \alpha_n \in \mathbf{R}_+, \quad (5.42)$$

co po porównaniu z (5.35) daje wartość stałej C równą:

$$C = 8 \sum_{n=1}^N \alpha_n^2. \quad (5.43)$$

Uwaga o stabilności rozwiązania: Warunek dodatniości pojemności informacyjnej otrzymany w (5.42), dla pojemności informacyjnej związanej z dodatnią częścią sygnatury metryki Minkowskiego, jest istotnym wynikiem z punktu widzenia teorii pomiaru. W rozważanym przykładzie analizy estymacyjnej wartości oczekiwanej energii cząstki gazu, został on otrzymany na gruncie samospójnego rozwiązania równań różniczkowych [49] informacyjnej obserwowanej zasady strukturalnej oraz zasady wariacyjnej. Niespełnienie tego warunku oznacza niestabilność badanego układu.

Z kolei, wstawiając otrzymaną wartość stałej C do (5.33) otrzymujemy:

$$Q(\Theta_\epsilon) = -4 \sum_{n=1}^N \alpha_n^2 < 0. \quad (5.44)$$

Natomiast sam problem wariacyjny (5.14), który można wyrazić po skorzystaniu z postaci $Q(\Theta_\epsilon)$ w (5.33) w postaci:

$$\delta_{(q_n)} \left(I(\Theta_\epsilon) + 4 \sum_{n=1}^N \alpha_n^2 - C \right) = 0, \quad (5.45)$$

jest ze swojej natury nieczuły na wartość stałej C .

Uwaga o randze amplitudy N : Istotnym zagadnieniem metody EFI jest wielkość próby N pobranej przez układ, tzn. ranga N amplitudy. Do sprawy liczby amplitud q_n wchodzących w opis układu podejmiemy w najprostszym z możliwych sposobów, sugerując stosowanie dwóch prostych i niewykluczających się kryteriów:

(1) Kryterium minimalizacji I ze względu na rangę amplitudy N [8]. Tzn. przy zachowaniu warunku $I > 0$, (2.225), ranga N może na tyle spaść, żeby istniało jeszcze samospójne rozwiązanie równań cząstkowych strukturalnej i wariacyjnej zasady informacyjnej.

Kryterium to nie oznacza nie realizowania rozwiązań z większą niż minimalna liczbą N .

(2) Kryterium obserwacyjne wyboru rangi amplitudy wiąże się z wyborem takiej wartości N , dla której otrzymane rozwiązanie ma realizację obserwowaną w eksperymencie. We współczesnych teoriach fizyki statystycznej oraz teoriach pola, realizowane są rozwiązania z niskimi wartościami rangi. Fakt ten zauważyliśmy już w Rozdziale 4 dla modeli teorii pola.

Rozważany w bieżącym rozdziale przykład rozkładu energii pozwala na ilustrację kryterium (1). W kolejnym z rozdziałów znajdziemy rozwiązanie EFI dla rozkładu prędkości cząsteczki gazu, stosując również kryterium (1). Obok stacjonarnego rozwiązania Maxwella-Boltzmanna z $N = 1$, wskażemy na fizyczną interpretację rozwiązań z $N > 1$.

Zastosowanie kryterium (1) dla rozkładu energii: Zauważmy, że ponieważ wszystkie α_n w (5.42) są rzeczywiste, zatem, przy założeniu braku wpływu nowych stopni swobody na poprzednie, pojemność informacyjna I wzrasta wraz ze wzrostem N . Zgodnie z kryterium (1), przyjmijmy dla rozważanego przypadku rozkładu fluktuacji energii, że:

$$N = 1. \quad (5.46)$$

Jedyny współczynnik α_1 oznaczmy teraz α , natomiast parametr $\Theta_\epsilon = (\theta_{\epsilon 1}) \equiv \theta_\epsilon$.

Zatem z (5.41) mamy amplitudę:

$$q(\mathbf{x}_\epsilon) = \sqrt{2\alpha} \exp [\alpha (\mathbf{x}_\epsilon^{\min} - \mathbf{x}_\epsilon)] , \quad (5.47)$$

której odpowiada rozkład gęstości prawdopodobieństwa fluktuacji energii \mathbf{x}_ϵ :

$$p(\mathbf{x}_\epsilon) = q^2(\mathbf{x}_\epsilon) = 2\alpha \exp [2\alpha (\mathbf{x}_\epsilon^{\min} - \mathbf{x}_\epsilon)] , \quad \alpha \in \mathbf{R}_+ . \quad (5.48)$$

Rozkład gęstości prawdopodobieństwa energii ϵ cząsteczki jest końcowym punktem analizy obecnego rozdziału. Ponieważ zgodnie z (5.1) mamy:

$$\mathbf{y}_\epsilon \equiv \frac{\epsilon}{c} = \theta_\epsilon + \mathbf{x}_\epsilon , \quad \frac{\epsilon_0}{c} \leq \mathbf{y}_\epsilon < \infty , \quad \text{dla} \quad \mathbf{x}_\epsilon^{\min} \leq \mathbf{x}_\epsilon < \infty , \quad (5.49)$$

gdzie $\epsilon_0/c = \theta_\epsilon + \mathbf{x}_\epsilon^{\min}$. Zatem $d\epsilon/d\mathbf{x}_\epsilon = c$, skąd rozkład dla zmiennej ϵ ma postać:

$$p(\epsilon) = p(\mathbf{x}_\epsilon) \frac{1}{|d\epsilon/d\mathbf{x}_\epsilon|} = 2\frac{\alpha}{c} \exp [-2\alpha (\epsilon - \epsilon_0)/c] , \quad \epsilon_0 \leq \epsilon < \infty . \quad (5.50)$$

Pozostało jeszcze określenie stałej α . Ponieważ wartość oczekiwana energii wynosi:

$$\langle E \rangle \equiv c \theta_\epsilon = \int_{\epsilon_0}^{+\infty} d\epsilon p(\epsilon) \epsilon, \quad (5.51)$$

więc wstawiając (5.50) do (5.51) otrzymuje się:

$$2\alpha = c(\langle E \rangle - \epsilon_0)^{-1}. \quad (5.52)$$

Zwróćmy uwagę, że z (5.51) wynika po pierwsze, że E jest estymatorem wartości oczekiwanej $\langle E \rangle$ energii cząstki:

$$\widehat{\langle E \rangle} = E, \quad (5.53)$$

a po drugie, że jest on nieobciążony.

Szukany rozkład gęstości prawdopodobieństwa energii cząstki ma zatem postać:

$$p(\epsilon) = \begin{cases} (\langle E \rangle - \epsilon_0)^{-1} \exp[-(\epsilon - \epsilon_0)/(\langle E \rangle - \epsilon_0)] & \text{dla } \epsilon \geq \epsilon_0 \\ 0 & \text{dla } \epsilon < \epsilon_0 \end{cases}, \quad (5.54)$$

gdzie w drugiej linii po prawej stronie zaznaczono fakt nie występowania cząstek z energią mniejszą niż ϵ_0 . Rozkład (5.54) jest końcowym rezultatem metody EFL. Jego postać daje zasadniczo rozkład Boltzmanna dla energii cząsteczki w gazie.

Rozkład Boltzmanna dla energii cząsteczki: Aby domknąć temat od strony fizycznej zauważmy, że energia całkowita cząsteczki wynosi $\epsilon = \epsilon_{kin} + V$, gdzie ϵ_{kin} jest energią kinetyczną cząsteczki a V jej energią potencjalną. Do potencjału V możemy dodać pewną stałą np. ϵ_0 , nie zmieniając przy tym fizycznego opisu zjawiska, zatem po przesunięciu ϵ o ϵ_0 otrzymujemy:

$$\epsilon_0 = 0 \quad \text{oraz} \quad \epsilon \geq 0. \quad (5.55)$$

Z kolei, dla cząstki gazu poruszającej się bez obrotu, *twierdzenie o ekwipartycji energii* mówi, że:

$$\langle E \rangle = \frac{3kT}{2}, \quad (5.56)$$

gdzie T jest temperaturą bezwzględną gazu. Wstawiając (5.55) wraz z (5.56) do (5.54) otrzymujemy:

$$p(\epsilon) = (3kT/2)^{-1} e^{-2\epsilon/3kT}, \quad \epsilon \geq 0, \quad (5.57)$$

czyli właściwą postać rozkładu Boltzmanna dla energii cząsteczki w gazie o temperaturze T .

Informacja Fishera dla θ_ϵ i DORC dla estymatora $\langle E \rangle$: W przypadku $N = 1$ oraz skalarne parametru θ_ϵ , pojemność informacyjna (5.42) jest równa informacji Fishera I_F dla tego parametru. Gdy dolne ograniczenie na energię cząstki wynosi $\epsilon_0 = 0$, otrzymujemy po skorzystaniu z (5.52) oraz (5.44) następujący związek:

$$I_F(\theta_\epsilon) = I(\theta_\epsilon) = -Q(\theta_\epsilon) = 4\alpha^2 = \frac{c^2}{\langle E \rangle^2} > 0. \quad (5.58)$$

Z Rozdziału 2.8.1 wiemy, że estymacja powyższego parametru oczekiwanego θ_ϵ , dualnego do 2α , dla standardowego rozkładu eksponentialnego, a takim jest rozkład Boltzmanna, spełnia DORC w Twierdzeniu Rao-Cramera. Zatem wariancja estymatora tego parametru wynosi:

$$\sigma^2(\hat{\theta}_\epsilon) = \frac{1}{I_F(\theta_\epsilon)} = \frac{\langle E \rangle^2}{c^2}, \quad (5.59)$$

skąd otrzymujemy wariancję estymatora (5.53) średniej energii cząstki, równą:

$$\sigma^2(\langle \widehat{E} \rangle) = \langle E \rangle^2. \quad (5.60)$$

Na tym kończymy analizę metody EFI dla rozkładu Boltzmanna.

Uwaga o równowadze statystycznej w metodzie EFI: Biorąc pod uwagę ograniczenia narzucone na normalizację (5.32) i skończoność wartości oczekiwanej (5.51) rozkładu, metoda EFI wyznacza rozkład przy narzuceniu zasad informacyjnych. Poprzez obserwowaną zasadę strukturalną dokonuje ona, dla gładkiej funkcji wiarygodności, separacji członu estymacyjnego dla parametru θ_ϵ związanego z gęstością pojemności i , której całka I jest śladem po metryce Rao-Fishera przestrzeni statystycznej \mathcal{S} , od członu strukturalnego q . Jest to wyrazem zasady Macha. Natomiast poprzez zasadę wariacyjną, metoda EFI dokonuje stabilizacji rozwiązania, wybierając najmniejszą odległość (liczoną wzdłuż geodezyjnej w przestrzeni statystycznej \mathcal{S}) wyestymowanego stanu układu od stanu zaobserwowanego. Geometria \mathcal{S} zależy od pojemności kanału informacyjnego I wyliczonej z uwzględnieniem jej zależności od informacji strukturalnej Q .

Uwaga: W związku z powyższym oraz poprzednio podanym kryterium (1) doboru rangi amplitudy N poprzez minimalizację I , pojawia się następujące, ogólne spojrzenie na estymacyjny charakter metody EFI.

Estymacja metodą EFI oznacza wybór równań generujących rozkład (lub równań ruchu) na skutek działania dwóch czynników. Pierwszy z nich wymaga, po pierwsze wzrostu informacji o układzie zawartej w gęstości pojemności informacyjnej \tilde{i}' określonej w (3.42) oraz (3.43), ale tylko na tyle na ile wymaga tego, zawarta w zasadzie strukturalnej $\tilde{i}' + \tilde{C} + \kappa q = 0$, (3.43), struktura układu opisana gęstością informacji strukturalnej q , (3.23), przy, po drugie jednoczesnej minimalizacji całkowitej fizycznej informacji $K = I + Q$, (3.31). Oszacowanie poszukiwanego równania następuje więc na skutek żądania samospójności rozwiązania dwóch różniczkowych zasad informacyjnych, (3.43) oraz (3.44).

Natomiast oczekiwana zasada strukturalna $I + \kappa Q = 0$, (3.33), jest w myśl związku (3.45) drugim, metrycznym czynnikiem, domykającym powyższą różniczkową analizę statystyczną doboru modelu. Dzieli ona modele statystyczne na dwie grupy: modele metryczne, równoważne zgodnie z (3.45) modelowi analitycznemu z metryką Rao-Fishera oraz modele, które nie spełniają zasady (3.45), czyli nierównoważne modelowi z metryką Rao-Fishera.

Związek EFI oraz ME dla standardowego rozkładu eksponentialnego: Poprzednio otrzymaliśmy rozkład Boltzmanna (2.258) jako szczególny przypadek rozwiązania zasady maksymalnej entropii (ME), realizowany w sytuacji warunku normalizacji oraz istnienia jednego parametru obserwowanego, którym jest średnia wartość energii cząstki. Szczególna postać standardowego rozkładu eksponentialnego oraz te same warunki brzegowe są przyczyną otrzymania tego samego rozwiązania w metodzie EFI oraz ME. W ogólności metoda EFI wykracza poza całą klasę modeli eksponentialnych metody ME.

5.1.4 Model Aoki-Yoshikawy dla produktywności branż

Model Aoki i Yoshikawy (AYM) został opracowany w celu opisu produktywności branż kraju [50, 51]. Rozważmy g ekonomicznych sektorów. Sektor i -ty jest scharakteryzowany przez *czynnik wielkości produkcji* n_i , tzn. liczebność, oraz *zmienną poziomu produktywności* A , tzn. wydajność jednostkową, przyjmującą wartości a_i .

Niech liczebność n_i będzie liczbą *aktywnych* pracowników w i -tym sektorze, co oznacza, że praca (robocizna) jest jedynym czynnikiem produkcyjnym. Zmienną losową w AYM jest poziom produktywności A , której rozkład jest określony poprzez parę (a_i, n_i) , $i = 1, 2, \dots, g$.

Unormowanie jako pierwszy warunek modelu: Załóżmy, że *całkowity zasób czynnika wielkości produkcji*, tzn. liczba dostępnych pracowników, jest w ekonomii zadany jako wielkość egzogeniczna, czyli nie kontrolowana od wewnątrz lecz zadana z zewnątrz. Niech jego wielkość jest równa n , co traktujemy jako *pierwsze ograniczenie* w modelu, tak, że zachodzi:

$$\sum_{i=1}^g n_i = n . \quad (5.61)$$

Sektory. Uporządkujmy wielkość produktywności a_i w sektorach od najmniejszej do największej:

$$a_1 < a_2 < \dots < a_g . \quad (5.62)$$

Ponieważ a_i jest poziomem produktywności i -tego sektora, zatem uzysk (wartość produkcji) w i -tym sektorze wynosi:

$$z_i = a_i n_i . \quad (5.63)$$

Zatem całkowity uzysk z w ekonomii kraju wynosi:

$$z = \sum_{i=1}^g z_i = \sum_{i=1}^g a_i n_i . \quad (5.64)$$

Drugi warunek modelu: Wielkość z jest interpretowana jako *produkt krajowy brutto* (PKB).

Założeniem AYM dla wartości PKB jest ustalenie wartości z poprzez egzogenicznie zadany agregatowy popyt D (demand), tzn.:

$$z = D . \quad (5.65)$$

Zatem *drugie ograniczenie* w modelu ma postać:

$$\sum_{i=1}^g a_i n_i = D . \quad (5.66)$$

Celem metody EFI dla modelu AYM jest, po pierwsze wyznaczenie równania generującego, a po drugie, teoretycznego rozkładu liczebności dla zmiennej poziomu produktywności A , tzn. określenie wektora obsadzeń:

$$\mathbf{n} = (n_1, n_2, \dots, n_g) . \quad (5.67)$$

Porównanie analizy dla AYM oraz rozkładu Boltzmann: Rozważane zagadnienie jest odpowiednikiem poprzedniego zagadnienia związanego z określeniem rozkładu prawdopodobieństwa energii

cząstki gazu. Można je określić jako zagadnienie rozmieszczenia n cząstek gazu na g poziomach energetycznych ϵ_i w warunkach równowagi statystycznej, w taki sposób, że zachowane są, liczba cząstek:

$$\sum_{i=1}^g n_i = n \quad (5.68)$$

oraz całkowita energia gazu \mathcal{E} :

$$\sum_{i=1}^g \epsilon_i n_i = \mathcal{E} \quad \text{lub} \quad \langle E \rangle = \sum_{i=1}^g \frac{n_i}{n} \epsilon_i = \frac{\mathcal{E}}{n} . \quad (5.69)$$

Zatem poziom produkcyjności a_i jest analogiem poziomu energetycznego ϵ_i , natomiast ograniczenie, które dał w AYM popyt D jest analogiem ograniczenia pochodzącego od wartości całkowitej energii \mathcal{E} gazu.

Dokonajmy następującego przyporządkowania pomiędzy wielkościami opisującymi rozkład energii cząstki gazu oraz rozkład poziomu produkcyjności. Po lewej stronie przyporządkowania “ \leftrightarrow ” jest wielkość dla energii cząstki, po prawej dla produkcyjności pracownika. Strzałki \rightarrow w nawiasach oznaczają przejście od ciągłego do dyskretnego rozkładu zmiennej (lub na odwrót).

Zatem, zmiennej energii cząstki E odpowiada poziom produkcyjności pracownika A :

$$E = (\epsilon \rightarrow \epsilon_i) \leftrightarrow A = (a_i \rightarrow a) . \quad (5.70)$$

Rozkładowi prawdopodobieństwa zmiennej energii cząstki $p(\epsilon)$ odpowiada rozkład poziomów produkcyjności pracownika $p(a)$:

$$\left(p(\epsilon) \rightarrow p_{\epsilon_i} = \frac{n_i}{n} \right) \leftrightarrow \left(p_i = \frac{n_i}{n} \rightarrow p(a) \right) . \quad (5.71)$$

Normalizacje rozkładów są sobie przyporządkowane następująco:

$$\left(\int_{\mathcal{Y}_\epsilon} d\epsilon p(\epsilon) = 1 \rightarrow \sum_{i=1}^g p_{\epsilon_i} = 1 \right) \leftrightarrow \left(\sum_{i=1}^g p_i = 1 \rightarrow \int_{\mathcal{Y}_a} da p(a) = 1 \right) . \quad (5.72)$$

Wartości oczekiwanej energii cząstki odpowiada wartość oczekiwana wartość produkcyjności pracownika:

$$\theta_\epsilon \equiv \langle E \rangle = \left(\int_{\mathcal{Y}_\epsilon} d\epsilon p(\epsilon) \epsilon \rightarrow \sum_{i=1}^g \frac{n_i}{n} \epsilon_i \right) \leftrightarrow \left(\sum_{i=1}^g p_i a_i \rightarrow \int_{\mathcal{Y}_a} da p(a) a \right) = \langle A \rangle \equiv \theta_A , \quad (5.73)$$

przy czym w AYM wartość oczekiwana produkcyjności jest zadana jako:

$$\langle A \rangle = D/n . \quad (5.74)$$

Określenie zmiennej addytywnych fluktuacji: Aby analiza w AYM mogła przebiegać dokładnie tak samo jak dla rozkładu Boltzmanna, musimy dokonać jeszcze jednego przejścia po stronie produkcyjności, a mianowicie przejść od poziomu produkcyjności A do jej fluktuacji X_a od wartości oczekiwanej $\langle A \rangle$, tzn. dokonać addytywnego rozkładu: $Y_a \equiv A = \langle A \rangle + X_a$.

Odpowiedni analog pomiędzy fluktuacjami energii i produktywności ma więc następującą postać: Dla energii cząstki zachodzi (5.1):

$$y_\epsilon = \frac{\epsilon}{c} = \theta_\epsilon + x_\epsilon, \quad \frac{\epsilon_0}{c} \leq y_\epsilon \leq \infty, \quad x_\epsilon^{min} \leq x_\epsilon < \infty, \quad (5.75)$$

gdzie skorzystano z założenia o nieograniczoności od góry fluktuacji energii (por. (5.39) wraz z dyskusją zawartą powyżej), natomiast dla produktywności pracownika zachodzi:

$$y_a \equiv a = \theta_A + x_a, \quad a_0 \leq y_a \leq \infty, \quad x_a^{min} = a_0 - \theta_A \leq x_a < \infty. \quad (5.76)$$

Równanie generujące amplitudę produktywności: Możemy teraz przenieść powyższą analizę EFI dla rozkładu Boltzmanna na grunt modelu AYM. Zatem wychodząc z zasady wariacyjnej (5.14) oraz strukturalnej (5.21) i odpowiednich analogów pojemności informacyjnej (5.11) oraz informacji strukturalnej (5.13), otrzymujemy zgodnie z analizą Rozdziału 5.1.3 równanie generujące rozkład produktywności (5.34), które dla wielkości próby $N = 1$ ma postać:

$$\frac{d^2 q(x_a)}{d x_a^2} = \alpha^2 q(x_a), \quad (5.77)$$

gdzie $q(x_a)$ jest amplitudą rozkładu fluktuacji produktywności, a α rzeczywistą stałą.

Rachunki analogiczne do przeprowadzonych pomiędzy (5.34) a (5.54), z warunkiem normalizacji (5.72), które poprzednio doprowadziły do amplitudy (5.47) rangi $N = 1$, dają następujące rozwiązanie równania (5.77) na amplitudę fluktuacji produktywności X_a w zakresie (5.76):

$$q(x_a) = \frac{1}{\sqrt{D/n}} \exp \left[- \frac{(D/n) - a_0 + x_a}{2 D/n} \right] \quad \text{dla} \quad x_a^{min} = a_0 - \theta_A \leq x_a < \infty \quad (5.78)$$

oraz w analogii do (5.54), rozkład gęstości produktywności A :

$$p(a) = \begin{cases} \frac{1}{(D/n) - a_0} \exp \left(- \frac{a - a_0}{(D/n) - a_0} \right) & \text{dla} \quad a \geq a_0 \\ 0 & \text{dla} \quad a < a_0 \end{cases}, \quad (5.79)$$

gdzie w drugiej linii po prawej stronie zaznaczono fakt nie występowania produktywności mniejszej niż a_0 . Rozkład (5.79) jest końcowym rezultatem metody EFI dla modelu AYM produktywności.

Porównanie wyników metody AYM oraz EFI: Aby postać rozkładu (5.79) oddać w pełni wynik Aoki i Yoshikawy⁵, należy powrócić do dyskretyzacji wartości $a \rightarrow a_i$ zmiennej A .

Rozwiązali oni powyższy problem stosując metodę mnożników Lagrange'a z warunkami ograniczającymi (5.61) oraz (5.66). Następnie założyli [50], że wartości produktywności są dyskretne, tworząc ciąg arytmetyczny:

$$a_i = i a_0 \quad \text{gdzie} \quad i = 1, 2, \dots, g, \quad (5.80)$$

gdzie a_0 jest najmniejszą produktywnością.

⁵Omówienie metody Aoki i Yoshikawy zamieszczono na końcu obecnego rozdziału.

Względny popyt agregatowy r : W końcu, przy założeniach, po pierwsze, że liczba dostępnych sektorów produkcyjności jest bardzo duża, tzn. $g \gg 1$ oraz po drugie, że:

$$r \equiv \frac{D/n}{a_0}, \quad (5.81)$$

tzn. *agregatowy popyt przypadający na jednego pracownika D/n odniesiony do najmniejszej produktywności a_0* , jest bardzo duży, Aoki i Yoshikawy otrzymali wynik [50, 51]:

$$P(i|\mathbf{n}^*) = \frac{n_i^*}{n} \approx \frac{1}{r-1} \left(\frac{r-1}{r} \right)^i \approx \left(\frac{1}{r} + \frac{1}{r^2} \right) e^{-\frac{i}{r}}, \quad i = 1, 2, \dots, \quad r \gg 1, \quad (5.82)$$

gdzie n_i^* , $i = 1, 2, \dots, g$, są współrzędnymi n_i wektora obsadzeń (5.67), przy których prawdopodobieństwo pojawienia się tego wektora obsadzeń jest maksymalne.

Wynik analizy metodą AYM: Rezultat (5.82) podaje *prawdopodobieństwo, że losowo wybrany pracownik jest w i -tym sektorze produkcyjności, o ile gospodarka znajduje się w stanie określonym wektorem obsadzeń $\mathbf{n}^* = (n_1^*, n_2^*, \dots, n_g^*)$* .

Przejście do rozkładu dyskretnego dla wyniku EFI: Aby porównać wynik (5.82) otrzymany w AYM z wynikiem (5.79) otrzymanym w metodzie EFI, przejdźmy w (5.79) do rozkładu dyskretnego. W tym celu musimy wyciągnąć wynik EFI w przedziale (a_i, a_{i+1}) , przy czym od razu założymy, że zachodzi $a_i = i a_0$, (5.80). W rezultacie otrzymujemy⁶:

$$P(i) = \int_{ia_0}^{(i+1)a_0} da p(a) = \left(1 - e^{-1/(r-1)} \right) e^{-(i-1)/(r-1)} \quad \text{dla} \quad i = 1, 2, \dots \quad (5.84)$$

Porównanie modeli dla $a_0 = 0$. Niech δa jest stałą szerokością sektorów produkcyjności. Wtedy, w przypadku gdy $a_0 = 0$, wzór (5.54) metody EFI prowadzi w miejsce (5.84) do rozkładu:

$$P(i) = \int_{(i-1)\delta a}^{i\delta a} da p(a) = \left(-1 + e^{1/\tilde{r}} \right) e^{-i/\tilde{r}}, \quad i = 1, 2, \dots \quad \text{dla} \quad a_0 = 0, \quad (5.85)$$

gdzie zamiast (5.81) wprowadziliśmy:

$$\tilde{r} \equiv \frac{D/n}{\delta a}, \quad (5.86)$$

jako *agregatowy popyt przypadający na jednego pracownika D/n odniesiony do szerokości sektora produkcyjności δa* . W granicy $\tilde{r} \gg 1$ z (5.85) otrzymujemy:

$$P(i) \approx \left(\frac{1}{\tilde{r}} + \frac{1}{2\tilde{r}^2} \right) e^{-i/\tilde{r}}, \quad i = 1, 2, \dots, \quad \text{dla} \quad a_0 = 0, \quad \tilde{r} \gg 1. \quad (5.87)$$

⁶ Przyjmując, że $r \gg 1$, otrzymujemy, w podanych granicach, następujący rozkład prawdopodobieństwa produktywności pracownika:

$$P(i) \approx \left(\frac{1}{r} + \frac{1}{2r^2} \right) \left(e^{-\frac{i}{r}} + \frac{1}{r} \right) \quad \text{dla} \quad i = 1, 2, \dots \quad \text{oraz} \quad a_0 > 0, \quad r \gg 1. \quad (5.83)$$

Wniosek z porównania wyników metody AYM oraz EFI: W granicy dużych \tilde{r} oba wyniki się schodzą. Jednakże granicę najmniejszej produktywności a_0 metoda EFI ujmuje inaczej niż AYM. To znaczy, formuła EFI (5.87) daje dla $a_0 = 0$ inną kwadratową poprawkę w \tilde{r} niż rezultat (5.82) modelu AYM dla $r = \tilde{r}$ oraz $a_0 \rightarrow 0$.

Uwaga: Ponadto wynik (5.85) jest dokładny, natomiast w AYM dyskretyzacja poziomu produktywności jest tylko wybiegiem technicznym, gdyż zmienna ta jest z natury ciągła, do czego i tak w końcu odwołuje się metoda AYM przy wyznaczaniu mnożników Lagrange’a, przechodząc z powodów rachunkowych w (5.61) oraz (5.66) z g do nieskończoności.

Uzupełnienie: Analiza Aoki i Yoshikawy dla produktywności. Podajmy sposób wyprowadzenia rozkładu wektora obsadzeń w AYM metodą czynników Lagrange’a [51]. Wprowadźmy wektor $\mathbf{H}^{(n)}$ indywidualnych przypisań, posiadający tyle składowych ilu jest pracowników w całej gospodarce kraju:

$$\mathbf{H}^{(n)} \equiv (H_1, H_2, \dots, H_n) = \mathbf{h} \equiv (h_1, h_2, \dots, h_n) . \quad (5.88)$$

Każda ze współrzędnych H_i , $i = 1, 2, \dots, n$, może przyjmować wartości $h_i = s$, gdzie $s \in \{1, 2, \dots, g\}$, co oznacza, że i -ty pracownik jest aktywny w s -tym sektorze gospodarki. Zatem wektor \mathbf{h} podaje jedną konfigurację indywidualnych przypisań. Jeśli ustalimy wektor obsadzeń \mathbf{n} , (5.67), to liczba W różnych \mathbf{h} (indywidualnych konfiguracji przypisań), realizujących ten sam ustalony wektor obsadzeń \mathbf{n} , wynosi [52]:

$$W(\mathbf{H}|\mathbf{n}) = \frac{n!}{\prod_{i=1}^g n_i!} . \quad (5.89)$$

Boltzman zauważył, że gdy układ znajduje się w równowadze statystycznej to, prawdopodobieństwo $\pi(\mathbf{n})$, że znajduje się on w stanie o określonym wektorze obsadzeń \mathbf{n} (tzn. że pojawił się taki właśnie wektor obsadzeń), jest proporcjonalne do liczby jego możliwych realizacji, tzn. do $W(\mathbf{H}|\mathbf{n})$. Zatem:

$$\pi(\mathbf{n}) = W(\mathbf{H}|\mathbf{n}) P(\mathbf{H}|\mathbf{n}) = \frac{n!}{\prod_{i=1}^g n_i!} \prod_{i=1}^g p^{n_i} = \frac{n!}{\prod_{i=1}^g n_i!} p^n = \mathcal{K} \frac{n!}{\prod_{i=1}^g n_i!} , \quad (5.90)$$

gdzie \mathcal{K} jest właściwą normalizacyjną stałą, a p jest *prawdopodobieństwem* zajęcia przez i -tego pracownika określonego s -tego sektora obsadzeń, *które zostało przyjęte jako takie samo dla wszystkich indywidualnych konfiguracji tych obsadzeń*. W celu rozwiązania postawionego problemu maksymalizacji prawdopodobieństwa $\pi(\mathbf{n})$ z warunkami (5.61) oraz (5.66), rozwiązujemy poniższy układ g równań:

$$\frac{\partial}{\partial n_i} \left[\ln \pi(\mathbf{n}) + \nu \left(\sum_{i=1}^g n_i - n \right) - \beta \left(\sum_{i=1}^g a_i n_i - D \right) \right] = 0 , \quad (5.91)$$

otrzymując jako rozwiązanie wektor obsadzeń \mathbf{n} :

$$n_i = n_i^* = e^\nu e^{-\beta a_i} , \quad i = 1, 2, \dots, g . \quad (5.92)$$

Stałe ν oraz β otrzymujemy wykorzystując (5.92) w (5.61) oraz (5.66).

Uwaga: Aby zapisać $\prod_{i=1}^g n_i!$ w formie nadającej się do minimalizacji, korzystamy z przybliżenia Stirling’a:

$$\ln \left[\prod_{i=1}^g n_i! \right] \approx \sum_{i=1}^g n_i (\ln n_i - 1) , \quad (5.93)$$

słusznego dla dużego układu z $n \gg 1$ oraz $n_i \gg 1$, $i = 1, 2, \dots, g$.

Układ w równowadze statystycznej: Przedstawiona w tym przypisie metoda znajdowania rozkładu w równowadze statystycznej zakłada spełnienie hipotezy wyrażonej równaniem (5.90) i mówiącej, że *wszystkie stany opisane wektorem indywidualnych przypisań $\mathbf{H}^{(n)}$, a spełniające warunki (5.61) oraz (5.66) są równie prawdopodobne*. Rozkład opisany wektorem \mathbf{n}^* jest w tym ujęciu sednem definicji rozkładu, będącego w równowadze statystycznej.

5.1.5 Rozkład Maxwella-Boltzmann dla prędkości

Poniżej wyznaczmy rozkład prędkości cząsteczki w gazie. Wychodząc z informacji Fishera (5.7) dla pędu oraz posługując się wariacyjną i strukturalną zasadą informacyjną (5.9) otrzymamy równanie generujące i znajdziemy jego rozwiązania, tzn. postać amplitud dla rozkładu prędkości.

Wartość współczynnika efektywności: Ze względu na spójność rozważań dla czterowektora pędu, przyjęcie współczynnika $\kappa = 1$ w rozważaniach dla energii skutkuje przyjęciem $\kappa = 1$ w analizie dla rozkładu pędu.

Pojemność informacyjna parametru $\Theta_{\vec{\varphi}}$ została podana w (5.7):

$$I(\Theta_{\vec{\varphi}}) = -4 \int_{\mathcal{X}_{\varphi}} d\vec{x}_{\varphi} \sum_{n=1}^N \sum_{k=1}^3 \left(\frac{\partial q_n(\vec{x}_{\varphi})}{\partial x_{\varphi_k}} \right)^2,$$

gdzie znaczenie znaku “minus” w definicji pojemności informacyjnej dla części pędowej zostało omówione w Rozdziale 5.1.2.

Informacja strukturalna i zależność \mathbf{q}_n od amplitudy oraz prędkości: Q dla parametru $\Theta_{\vec{\varphi}}$ ma postać:

$$Q(\Theta_{\vec{\varphi}}) = \int_{\mathcal{X}_{\varphi}} d\vec{x}_{\varphi} \sum_{n=1}^N q_n^2(\vec{x}_{\varphi}) \mathbf{q}_n(q_n(\vec{x}_{\varphi}), \vec{x}_{\varphi}). \quad (5.94)$$

Wprowadzenie do obserwowanej informacji strukturalnej $\mathbf{q}_n(q_n(\vec{x}_{\varphi}), \vec{x}_{\varphi})$ jej jawnej zależności nie tylko od amplitud q_n , ale również od pędu \vec{x}_{φ} , pozwala na rozważenie szerszego zakresu zagadnień niż to miało miejsce dla przypadku rozkładu energii, mianowicie pozwala na rozważenie rozwiązań *nierównowagowych*.

Obserwowana, strukturalna zasada informacyjna: Podobnie jak to uczyniliśmy w (5.18) dla energii, tak i teraz dla pojemności informacyjnej $I(\Theta_{\vec{\varphi}})$, (5.7), dokonamy całkowania przez części. Zakładając dodatkowo, że amplitudy dla prędkości $q_n(\vec{x}_{\varphi})$ znikają w \pm nieskończoności, **oczekiwana strukturalna zasada informacyjna** $\tilde{i}'(\Theta_{\vec{\varphi}}) + \tilde{\mathbf{C}}_{\vec{\varphi}} + \kappa q(\Theta_{\vec{\varphi}}) = 0$ w (5.9) przyjmuje postać:

$$\tilde{i}'(\Theta_{\vec{\varphi}}) + q(\Theta_{\vec{\varphi}}) = 4 \sum_{n=1}^N \left[q_n(\vec{x}_{\varphi}) \sum_{k=1}^3 \frac{\partial^2 q_n(\vec{x}_{\varphi})}{\partial x_{\varphi_k}^2} + \frac{1}{4} q_n^2(\vec{x}_{\varphi}) \mathbf{q}_n(q_n(\vec{x}_{\varphi}), \vec{x}_{\varphi}) \right] = 0, \quad (5.95)$$

gdzie fakt, że $\tilde{\mathbf{C}}_{\vec{\varphi}} = 0$, nie wnosząc tym samym wkładu w powyższą zasadę strukturalną, wynika ze znikania amplitud w nieskończoności.

Wariacyjna zasada informacyjna w (5.9) przyjmuje postać:

$$\delta_{(q_n)}(I(\Theta_{\vec{\varphi}}) + Q(\Theta_{\vec{\varphi}})) = 4 \int_{\mathcal{X}_{\varphi}} d\vec{x}_{\varphi} \sum_{n=1}^N \left[- \sum_{k=1}^3 \left(\frac{\partial q_n(\vec{x}_{\varphi})}{\partial x_{\varphi_k}} \right)^2 + \frac{1}{4} q_n^2(\vec{x}_{\varphi}) \mathbf{q}_n(q_n(\vec{x}_{\varphi}), \vec{x}_{\varphi}) \right] = 0. \quad (5.96)$$

Równania Eulera-Lagrange’a: Rozwiązaniem N -funkcyjnego problemu wariacyjnego (5.96) jest układ

równań Eulera-Lagrange'a:

$$\sum_{m=1}^3 \frac{\partial}{\partial x_{\varphi m}} \left(\frac{\partial k(q_n(\vec{x}_\varphi), \vec{x}_\varphi)}{\partial \left(\frac{\partial q_n(\vec{x}_\varphi)}{\partial x_{\varphi m}} \right)} \right) = \frac{\partial k(q_n(\vec{x}_\varphi), \vec{x}_\varphi)}{\partial q_n(\vec{x}_\varphi)}, \quad n = 1, 2, \dots, N, \quad (5.97)$$

gdzie

$$k(q_n(\vec{x}_\varphi), \vec{x}_\varphi) = 4 \left[- \sum_{k=1}^3 \left(\frac{\partial q_n(\vec{x}_\varphi)}{\partial x_{\varphi k}} \right)^2 + \frac{1}{4} q_n^2(\vec{x}_\varphi) \mathfrak{F}_n(q_n(\vec{x}_\varphi), \vec{x}_\varphi) \right] \quad (5.98)$$

jest gęstością informacji fizycznej (4.7) zdefiniowaną w Rozdziale 4.2.

Niejemność k : Zauważmy, że z (5.98) i z żądania niejemności informacji fizycznej k na poziomie obserwowanym, wynika:

$$\mathfrak{F}_n(q_n(\vec{x}_\varphi), \vec{x}_\varphi) \geq 0, \quad (5.99)$$

tzn. *niejemność obserwowanej informacji strukturalnej w analizie estymacyjnej wartości oczekiwanej pędu cząsteczki gazu.*

Uwaga: Poniżej, ze względu na uproszczenie zapisu, pominiemy zaznaczenie fluktuacji pędu \vec{x}_φ w argumentie amplitudy $q_n(\vec{x}_\varphi)$.

Układ równań różniczkowych: Dla każdego $n = 1, 2, \dots, N$, zasada strukturalna (5.95) jest na poziomie obserwowanym następująca:

$$q_n \sum_{m=1}^3 \frac{\partial^2 q_n}{\partial x_{\varphi m}^2} + \frac{1}{4} q_n^2 \mathfrak{F}_n(q_n, \vec{x}_\varphi) = 0. \quad (5.100)$$

Natomiast każde z N równań Eulera-Lagrange'a (5.97) ma postać następującego równania różniczkowego:

$$\sum_{m=1}^3 \frac{\partial^2 q_n}{\partial x_{\varphi m}^2} = - \frac{1}{2} \frac{\partial \left(\frac{1}{4} q_n^2 \mathfrak{F}_n(q_n, \vec{x}_\varphi) \right)}{\partial q_n}. \quad (5.101)$$

Równanie strukturalne (5.100) wraz z równaniem Eulera-Lagrange'a (5.101) posłuży do wyprowadzenia równania generującego rozkład Maxwella-Boltzamnna.

Wyprowadzenie równania generującego: Równania (5.100) oraz (5.101) pozwalają wyeliminować występującą w nich sumę, dając równanie:

$$\frac{1}{2} \frac{\partial \left(\frac{1}{4} q_n^2 \mathfrak{F}_n(q_n, \vec{x}_\varphi) \right)}{\partial q_n} = \frac{\frac{1}{4} q_n^2 \mathfrak{F}_n(q_n, \vec{x}_\varphi)}{q_n}, \quad (5.102)$$

które po obustronnym scałkowaniu prowadzi do rozwiązania:

$$\frac{1}{4} q_n^2 \mathfrak{F}_n(q_n, \vec{x}_\varphi) = q_n^2 f_n(\vec{x}_\varphi), \quad (5.103)$$

lub

$$\mathfrak{F}_n(q_n, \vec{x}_\varphi) = 4 f_n(\vec{x}_\varphi) \geq 0. \quad (5.104)$$

Funkcja $f_n(\vec{x}_\varphi)$ nie zależy od amplitudy q_n i pojawiła się w wyniku całkowania równania (5.102) jako pewna stała całkowania w znaczeniu jej niezależności od amplitudy q_n . Zaznaczona nieujemność funkcji $f_n(\vec{x}_\varphi)$ w jej dziedzinie wynika z nieujemności $\Phi_n(q_n, \vec{x}_\varphi)$ otrzymanej w (5.99).

Równanie generujące: Wykorzystując (5.103) w (5.100) eliminujemy obserwowaną informację strukturalną, otrzymując *równanie generujące* dla amplitudy rozkładu (fluktuacji) prędkości \vec{x}_φ cząsteczki w gazie:

$$\nabla^2 q_n(\vec{x}_\varphi) = -q_n(\vec{x}_\varphi) f_n(\vec{x}_\varphi) , \quad (5.105)$$

gdzie $\nabla^2 \equiv \sum_{m=1}^3 \partial^2 / \partial x_{\varphi m}^2$ jest operatorem Laplace'a ze względu na współrzędne pędowe $x_{\varphi m}$.

Równanie (5.105), podobnie jak (5.34), pojawiło się jako rozwiązanie strukturalnej i wariacyjnej zasady informacyjnej.

Rozwiązanie q_n równania (5.105) jest *samospójnym* rozwiązaniem sprzężonego układu równań różniczkowych (5.100) i (5.101), utworzonych przez parę zasad informacyjnych.

Rozwiązanie równania generującego: Poniżej rozwiążemy równanie generujące (5.105), czyniąc kilka fizycznych założeń odnośnie postaci funkcji $f_n(\vec{x}_\varphi)$. Jej postać wpływa na postać otrzymanych amplitud $q_n(\vec{x}_\varphi)$, a zatem również na rozkład $p(\vec{x}_\varphi)$.

Fizyczne założenia o postaci $f_n(\vec{x}_\varphi)$: Załóżmy, że każda gęstość rozkładu prawdopodobieństwa $p_n(\vec{x})$ jest taką samą funkcją parzystą każdej współrzędnej $x_{\varphi i}$ (fluktuacji) pędu \vec{x}_φ . W konsekwencji *układ jest izotropowy*, tzn. rozkład prawdopodobieństwa dla pędu nie zależy od kierunku w bazowej przestrzeni położań.

Następnie zakładamy *nierelatywistyczne przybliżenie*, co oznacza, że prędkości cząsteczek są dużo mniejsze od prędkości światła c , czyli fluktuacja pędu cząsteczki $x_{\varphi i}$ jest również mała w porównaniu z mc .

Z powyższych założeń wynika ogólna postać funkcji $f_n(\vec{x}_\varphi)$. Otóż jej rozwinięcie w szereg potęgowy ma tylko składowe parzyste modułu fluktuacji pędu $|\vec{x}_\varphi|$, a ponieważ wartość $|\vec{x}_\varphi|$ jest mała, zatem szereg ten obetniemy na drugim wyrazie:

$$f_n(\vec{x}_\varphi) = A_n + B |\vec{x}_\varphi|^2 , \quad A_n, B = const. \quad (5.106)$$

Równanie generujące z f_n : Podstawiając (5.106) do (5.105), otrzymujemy następującą postać równania generującego:

$$\nabla^2 q_n(\vec{x}_\varphi) + (A_n + B |\vec{x}_\varphi|^2) q_n(\vec{x}_\varphi) = 0 \quad \text{dla } n = 1, 2, \dots, N . \quad (5.107)$$

Przejdźmy do nowego indeksu:

$$n' := n - 1 = 0, 1, \dots, N - 1 . \quad (5.108)$$

Wtedy w miejsce (5.107) otrzymujemy równanie:

$$\nabla^2 q_{n'}(\vec{x}_\varphi) + (A_{n'} + B |\vec{x}_\varphi|^2) q_{n'}(\vec{x}_\varphi) = 0 \quad \text{dla } n' = 1, 2, \dots, N - 1 , \quad (5.109)$$

które po dokonaniu separacji zmiennych kartezjańskich i faktoryzacji amplitudy:

$$q_{n'}(\vec{x}_\varphi) = q_{n'_1}(x_{\varphi 1}) q_{n'_2}(x_{\varphi 2}) q_{n'_3}(x_{\varphi 3}) , \quad (5.110)$$

przechodzi w równoważny mu układ trzech równań różniczkowych:

$$q_{n'_i}''(x_{\varphi_i}) + \left(A_{n'_i} + B x_{\varphi_i}^2\right) q_{n'_i}(x_{\varphi_i}) = 0, \quad i = 1, 2, 3, \quad \sum_{i=1}^3 A_{n'_i} \equiv A_{n'} . \quad (5.111)$$

Gdy stałe równania (5.111) mają postać:

$$A_{n'_i} = \frac{n'_i + 1/2}{a_0^2}, \quad B = -\frac{1}{4a_0^4}, \quad a_0 = \text{const.}, \quad n'_i = 0, 1, \dots, \quad (5.112)$$

wtedy ma ono rozwiązanie.

Postać rozwiązania: Ponieważ każde z równań (5.111) jest równaniem Helmholtz'a, zatem jego rozwiązaniami są paraboliczno-cylindryczne funkcje [53, 54]:

$$q_{n'_i}(x_{\varphi_i}) = e^{-x_{\varphi_i}^2/(4a_0^2)} 2^{-n'_i/2} H_{n'_i}\left(\frac{x_{\varphi_i}}{a_0\sqrt{2}}\right), \quad \text{gdzie } n'_i = 0, 1, \dots, \quad i = 1, 2, 3, \quad (5.113)$$

gdzie $H_{n'_i}$ są wielomianami Hermite'a:

$$H_j(t) = j! \sum_{m=0}^{[j/2]} (-1)^m \frac{(2t)^{j-2m}}{m!(j-2m)!}, \quad j = 0, 1, \dots, \quad (5.114)$$

gdzie $[j/2]$ oznacza liczbę całkowitą nie przekraczającą $j/2$.

Amplitudy fluktuacji pędu: Wstawiając (5.113) do (5.110) otrzymujemy szukaną postać amplitudy fluktuacji pędu:

$$q_{n'}(\vec{x}_{\varphi}) = e^{-|\vec{x}_{\varphi}|^2/(4a_0^2)} 2^{-n'/2} \cdot \sum_{\substack{ijk \\ i+j+k=n'}} a_{n'_{ijk}} H_i\left(\frac{x_{\varphi_1}}{a_0\sqrt{2}}\right) H_j\left(\frac{x_{\varphi_2}}{a_0\sqrt{2}}\right) H_k\left(\frac{x_{\varphi_3}}{a_0\sqrt{2}}\right), \quad \text{gdzie } a_{n'_{ijk}} = \text{const.}, \quad (5.115)$$

przy czym:

$$\sum_{i=1}^3 n'_i = n' = 0, 1, \dots, N-1. \quad (5.116)$$

Rozkład fluktuacji pędu: Funkcja rozkładu prawdopodobieństwa ma więc zgodnie z (3.61) postać:

$$p(\vec{x}_{\varphi}) = \frac{1}{N} \sum_{n'=0}^{N-1} q_{n'}^2(\vec{x}_{\varphi}) = p_0 e^{-|\vec{x}_{\varphi}|^2/(2a_0^2)} \cdot \left\{ 1 + \sum_{n'=1}^{N-1} 2^{-n'} \left[\sum_{\substack{ijk \\ i+j+k=n'}} b_{n'_{ijk}} H_i\left(\frac{x_{\varphi_1}}{a_0\sqrt{2}}\right) H_j\left(\frac{x_{\varphi_2}}{a_0\sqrt{2}}\right) H_k\left(\frac{x_{\varphi_3}}{a_0\sqrt{2}}\right) \right]^2 \right\}, \quad (5.117)$$

gdzie $p_0 = a_{000}^2/N$, a liczba 1 w nawiasie klamrowym pochodzi z $n' = 0$ w sumie w pierwszej równości, natomiast nowe stałe $b_{n'_{ijk}}$ są proporcjonalne do stałych $a_{n'_{ijk}}$.

Po raz pierwszy takie wyprowadzenie postaci funkcji rozkładu gęstości prawdopodobieństwa dla fluktuacji pędu zostało podane w [8].

Rozwiązania równowagowe i nierównowagowe: Wartość $N = 1$, dla której (5.117) jest rozkładem Gaussa, daje równowagowy rozkład Maxwella-Boltzmannna dla (fluktuacji) pędu cząsteczki gazu. Pozostałe rozwiązania dla $N \geq 2$ dają rozwiązania nierównowagowe [8]. Jednak i one są stacjonarne ze względu na fakt bycia samospójnymi rozwiązaniami układu sprzężonych równań różniczkowych zasady strukturalnej (5.100) i wariacyjnej (5.101).

Wcześniej, rozkłady (5.117) z $N \geq 2$ zostały odkryte przez Rumer'a i Ryskin'a [55] jako rozwiązania równania transportu Boltzmann'a.

Interferencja rozwiązań: Zwróćmy uwagę, że rozwiązanie (5.117) implikuje interferencję pomiędzy wyrażeniami iloczynowymi występującymi w amplitudach (5.115). Interferencja pojawiła się więc jako cecha charakterystyczna dla rozwiązywanego równania różniczkowego oraz wprowadzenia amplitud do opisu układu, a nie jako cecha charakterystyczna wyłącznie mechaniki kwantowej.

Rozkład Maxwella-Boltzmannna dla prędkości z $N = 1$: Założyliśmy na wstępie, że zbiornik zawierający cząsteczki gazu jest w spoczynku, w pewnym inercjalnym układzie współrzędnych. Zatem średnia prędkość cząsteczki wynosi zero, tzn. $\vec{\theta}_\varphi = 0$ i z (5.2) otrzymujemy $\vec{y}_\varphi = \vec{x}_\varphi = \vec{\varphi}$.

Tak więc, dla $N = 1$ rozkład gęstości prawdopodobieństwa pędu (5.117) przyjmuje postać:

$$p(\vec{\varphi}) = p_0 e^{-|\vec{\varphi}|^2/(2a_0^2)} = p_0 e^{-\varphi_1^2/(2a_0^2)} e^{-\varphi_2^2/(2a_0^2)} e^{-\varphi_3^2/(2a_0^2)}, \quad (5.118)$$

gdzie $\vec{\varphi} = (\varphi_1, \varphi_2, \varphi_3)$ oraz $-\infty < \varphi_i < \infty$, $i = 1, 2, 3$. Wyznamy wartość stałej a_0 dla przypadku zerowej energii potencjalnej. Z warunku normalizacji $\int d\vec{\varphi} p(\vec{\varphi}) = 1$ wyznaczamy, że stała p_0 w (5.118) wynosi:

$$p_0 = \frac{1}{(2\pi)^{3/2} a_0^3}. \quad (5.119)$$

Ponieważ wtedy $\langle E \rangle = \langle \vec{\varphi}^2 \rangle / (2m)$, a z twierdzenia o ekwipartycji energii wiemy, że $\langle E \rangle = 3kT/2$, zatem

$$\langle \vec{\varphi}^2 \rangle = 3m kT. \quad (5.120)$$

Mając wartość p_0 , wyznaczmy wartość oczekiwaną:

$$\langle \vec{\varphi}^2 \rangle = \int d\vec{\varphi} p(\vec{\varphi}) \vec{\varphi}^2 \quad (5.121)$$

i przyrównajmy ją do (5.120). W rezultacie otrzymujemy:

$$a_0^2 = mc^2 kT. \quad (5.122)$$

Rozkładu wartości pędu: Kolejnym krokiem jest wyznaczenie rozkładu prawdopodobieństwa wartości pędu $\varphi = |\vec{\varphi}|$. Przechodząc od współrzędnych kartezjańskich pędu $(\varphi_1, \varphi_2, \varphi_3)$ do współrzędnych sferycznych (φ, θ, ϕ) , otrzymujemy:

$$p(\varphi, \theta, \phi) = |J| p(\varphi_1, \varphi_2, \varphi_3), \quad J = \varphi^2 \sin \theta, \quad (5.123)$$

gdzie J jest jacobianem przejścia. Wstawiając teraz (5.119) wraz z (5.122) do (5.118) otrzymujemy rozkład $p(\wp_1, \wp_2, \wp_3)$ ze znanymi już stałymi a_0 oraz p_0 . Wynik ten wstawiając do (5.123) i wycalkowując po zmiennych θ oraz ϕ , otrzymujemy szukany rozkład gęstości prawdopodobieństwa dla wartości pędu:

$$p(\wp) = \sqrt{2/\pi} (mkT)^{-3/2} \wp^2 e^{-\wp^2/2mkT} . \quad (5.124)$$

Prawo Maxwella-Boltzmanna: Rozkład (5.124) wyraża tzw. prawo Maxwella-Boltzmanna, a podstawiając $\wp = mv$ w (5.124), w miejsce wartości pędu, otrzymujemy rozkład Maxwella-Boltzmanna dla wartości prędkości v cząsteczki gazu.

5.1.6 Informacja Fishera jako ograniczenie dla wzrostu entropii

Rozważmy układ zawierający *jedną* lub *więcej* cząstek poruszających się w sposób losowy, w pewnym zamkniętym obszarze. Niech rozważany obszar będzie izolowany, tzn. żadna cząstka ani z obszaru nie ucieka ani do niego nie przechodzi. Brzeg (\vec{B}) obszaru jest zadany wektorem wodzącym $\vec{b} \in \vec{B}$. W pewnej chwili układ dokonuje pomiaru jego czterowektora położenia t, \vec{y} , zgodnie z rozkładem gęstości prawdopodobieństwa $p(\vec{y}, t)$ ⁷.

Niech (y^1, y^2, y^3) są współrzędnymi kartezjańskimi wektora \vec{y} **położenia przestrzennego**, a $r = |\vec{y}|$ jego długością. Ponieważ założyliśmy, że cząstka znajduje się gdzieś wewnątrz obszaru więc:

$$p(t) = \int d\vec{y} p(\vec{y}, t) = 1 . \quad (5.125)$$

Oznaczmy przez $S_H(t)$ entropię Shannona układu (2.147) w chwili czasu t :

$$S_H(t) = - \int d\vec{y} p(\vec{y}, t) \ln p(\vec{y}, t) . \quad (5.126)$$

Drugą zasadą termodynamiki: Można pokazać (por. Dodatek 7.4), że entropia Shannona spełnia drugą zasadę termodynamiki:

$$\frac{dS_H(t)}{dt} \geq 0 , \quad (5.127)$$

która określa *dolną granicę tempa zmiany entropii*. Poniższe rozważania poświęcone są znalezieniu ograniczenia na jego górną granicę.

Równanie ciągłości strumienia: Ponieważ żadna cząstka nie opuszcza ani nie wpływa do obszaru, zatem spełnione jest równanie ciągłości strumienia prawdopodobieństwa:

$$\frac{\partial p(\vec{y}, t)}{\partial t} + \vec{\nabla} \cdot \vec{P}(\vec{y}, t) = 0 , \quad (5.128)$$

gdzie $\vec{P}(\vec{y}, t)$ jest *prądem prawdopodobieństwa*, którego konkretna postać zależy od rozważanego układu.

Warunki brzegowe: Z kolei określmy warunki brzegowe spełniane przez gęstość prawdopodobieństwa i jego prąd. Będą nam one przydatne dla rozwiązania postawionego sobie zadania.

⁷ Jeśli układ opisany jest funkcją falową $\psi(\vec{y}, t)$, wtedy $p(\vec{y}, t) = |\psi(\vec{y}, t)|^2$.

Ponieważ żadne cząstki nie przechodzą przez granicę obszaru, zatem \vec{P} spełnia warunek brzegowy Dirichleta:

$$\vec{P}(\vec{y}, t) \Big|_{\vec{y} \in \vec{B}} = 0 . \quad (5.129)$$

Założmy dodatkowo, że jeśli brzeg obszaru znajduje się w nieskończoności, to:

$$\lim_{\vec{y} \rightarrow \infty} \vec{P}(\vec{y}, t) \rightarrow 0 , \quad \text{szybciej niż } 1/r^2 . \quad (5.130)$$

Ponieważ obszar jest odizolowany, zatem prawdopodobieństwo, że cząstka znajduje się na jego granicy znika, co oznacza, że $p(\vec{y}, t)$ również spełnia warunek brzegowy Dirichleta:

$$p(\vec{y}, t) \Big|_{\vec{y} \in \vec{B}} = 0 . \quad (5.131)$$

Ze względu na normalizację:

$$\int d\vec{y} p(\vec{y}, t) = 1 , \quad (5.132)$$

zakładamy, że w przypadku brzegu nieskończenie oddalonego, $p(\vec{y}, t)$ spełnia warunek:

$$\lim_{\vec{y} \rightarrow \infty} p(\vec{y} | t) \rightarrow 0 , \quad \text{szybciej niż } 1/r^3 . \quad (5.133)$$

W końcu dla domknięcia potrzebnych warunków brzegowych przyjmujemy:

$$\vec{P} \ln p \Big|_{\vec{y} \in \vec{B}} = 0 . \quad (5.134)$$

Wyprowadzenie ograniczenia na tempo wzrostu entropii: Przystąpmy teraz do sedna rachunków. Różniczkowanie (5.126) po $\frac{\partial}{\partial t}$ daje:

$$\frac{\partial S_H}{\partial t} = - \frac{\partial}{\partial t} \int d\vec{y} p \ln p = - \int d\vec{y} \frac{\partial p}{\partial t} \ln p - \int d\vec{y} p \frac{1}{p} \frac{\partial p}{\partial t} . \quad (5.135)$$

Druga całka po prawej stronie daje po skorzystaniu z warunku unormowania:

$$\frac{\partial}{\partial t} \int d\vec{y} p = 0 . \quad (5.136)$$

Oznaczmy przez (P^1, P^2, P^3) kartezjańskie składowe prądu \vec{P} . Podstawiając (5.128) do pierwszej całki po prawej stronie w (5.135) otrzymujemy:

$$\frac{\partial S_H}{\partial t} = \int d\vec{y} \vec{\nabla} \cdot \vec{P} \ln p = \int \int \int dy^3 dy^2 dy^1 \left[\frac{\partial}{\partial y^1} P^1 + \frac{\partial}{\partial y^2} P^2 + \frac{\partial}{\partial y^3} P^3 \right] \ln p , \quad (5.137)$$

gdzie $\nabla_i = \frac{\partial}{\partial y^i}$.

Całkując przez części wewnętrzne całki $\int dy^i \frac{\partial}{\partial y^i} P^i$ w (5.137) dla trzech składników $i = 1, 2, 3$, otrzymujemy dla każdego z nich:

$$\int dy^i \left(\frac{\partial}{\partial y^i} P^i \right) \ln p = P^i \ln p \Big|_{\vec{y} \in \vec{B}} - \int dy^i \frac{P^i}{p} \frac{\partial p}{\partial y^i} = - \int dy^i P^i \frac{\partial p}{\partial y^i} \frac{1}{p} , \quad i = 1, 2, 3 , \quad (5.138)$$

gdzie skorzystano z (5.134). Zatem:

$$\frac{\partial S_H}{\partial t} = - \int dy^3 dy^2 dy^1 \left(P^1 \frac{\partial p}{\partial y^1} + P^2 \frac{\partial p}{\partial y^2} + P^3 \frac{\partial p}{\partial y^3} \right) \frac{1}{p} = - \int d\vec{y} \left(\vec{P} \cdot \vec{\nabla} p \right) \frac{1}{p} . \quad (5.139)$$

Z powyższego, po prostych przekształceniach mamy:

$$\left(\frac{\partial S_H}{\partial t}\right)^2 = \left[\int d\vec{y} \left(\frac{\vec{P}}{\sqrt{p}} \right) \cdot \left(\frac{\sqrt{p} \vec{\nabla} p}{p} \right) \right]^2 = \left[\sum_{i=1}^3 \int d\vec{y} \left(\frac{P^i}{\sqrt{p}} \right) \left(\frac{\sqrt{p} \nabla_i p}{p} \right) \right]^2. \quad (5.140)$$

Z nierówności Schwartza otrzymujemy:

$$\left[\sum_{i=1}^3 \int d\vec{y} \left(\frac{P^i}{\sqrt{p}} \right) \left(\frac{\sqrt{p} \nabla_i p}{p} \right) \right]^2 \leq \left[\sum_{i=1}^3 \int d\vec{y} \left(\frac{P^i}{\sqrt{p}} \right)^2 \right] \left[\sum_{i=1}^3 \int d\vec{y} \left(\frac{\sqrt{p} \nabla_i p}{p} \right)^2 \right]. \quad (5.141)$$

Ogólna postać ograniczenia na tempo wzrostu entropii: Ostatecznie z (5.140) i (5.141) oraz po zastąpieniu sumowań po i znakiem iloczynów skalarnych, dostajemy:

$$\left(\frac{\partial S_H}{\partial t}\right)^2 \leq \int d\vec{y} \frac{\vec{P} \cdot \vec{P}}{p} \int d\vec{y} \frac{\vec{\nabla} p \cdot \vec{\nabla} p}{p}. \quad (5.142)$$

Przejście do postaci z pojemnością informacyjną: Ponieważ gradient $\vec{\nabla} \equiv (\frac{\partial}{\partial y^1}, \frac{\partial}{\partial y^2}, \frac{\partial}{\partial y^3})$ zawiera różniczkowanie po wartościach pomiarowych y^i , a nie fluktuacjach x^i , jak to jest w kinematycznej postaci pojemności informacyjnej Friedena-Soffera, zatem musimy uczynić dodatkowe założenie.

Założenie niezmienniczości rozkładu ze względu na przesunięcie: Przejdźmy do addytywnych przesunięć, $\vec{x} = \vec{y} - \vec{\theta}$. Ponieważ $\vec{x} = (x^i)_{i=1}^3$ oraz $\vec{y} = (y^i)_{i=1}^3$ różnią się o stałą wartość oczekiwaną położenia $\vec{\theta} = (\theta^i)_{i=1}^3$, zatem zakładając dodatkowo, że rozkład p jest niezmienniczy ze względu na przesunięcie $\vec{\theta}$, otrzymujemy:

$$\frac{\partial p}{\partial y^i} = \frac{\partial p}{\partial x^i}, \quad i = 1, 2, 3. \quad (5.143)$$

Przy powyższym założeniu, zachodzi:

$$\int d\vec{y} \frac{\vec{\nabla} p(\vec{y}, t) \cdot \vec{\nabla} p(\vec{y}, t)}{p(\vec{y}, t)} = \int d\vec{x} \frac{\vec{\nabla} p(\vec{x}, t) \cdot \vec{\nabla} p(\vec{x}, t)}{p(\vec{x}, t)} \quad (5.144)$$

i okazuje się, że druga całka po prawej stronie nierówności w (5.142) jest pojemnością informacyjną trzech kanałów przestrzennych dla $N = 1$ w chwili czasu t :

$$I \equiv I(t) = \int d\vec{x} \frac{\nabla p(\vec{x}, t) \cdot \nabla p(\vec{x}, t)}{p(\vec{x}, t)}. \quad (5.145)$$

Ograniczenie na tempo wzrostu S_H układu z niezmienniczością przesunięcia: Korzystając z powyższej postaci pojemności informacyjnej, możemy zapisać (5.142) następująco:

$$\left(\frac{\partial S_H}{\partial t}\right)^2 \leq I(t) \int d\vec{x} \frac{\vec{P} \cdot \vec{P}}{p}, \quad \text{lub} \quad \left(\frac{\partial S_H}{\partial t}\right) \leq \sqrt{I(t)} \sqrt{\int d\vec{x} \frac{\vec{P} \cdot \vec{P}}{p}}, \quad (5.146)$$

Pokazaliśmy zatem, że przy założeniu niezmienniczości rozkładu za względu na przesunięcie, tempo wzrostu entropii jest obustronnie ograniczone. Ograniczenie dolne (5.127) wyraża zasadę nie-malenia entropii Shannona w czasie. Jej termodynamicznym odpowiednikiem jest twierdzenie H Boltzmanna.

Wniosek: Nierówność (5.146) oznacza, że ograniczenie górne tempa wzrostu entropii jest proporcjonalne do pierwiastka z pojemności informacyjnej (5.145) dla pomiaru położenia \vec{y} . Jest to jeden z nowych wyników teorii pomiaru otrzymany przez Friedena, Soffera, Plastino i Plastino [8]. Jego termodynamiczne konsekwencje czekają na weryfikację.

W [8] podano przykłady zastosowania tego twierdzenia dla strumienia cząstek klasycznych, strumienia w elektrodynamice klasycznej i strumienia cząstek ze spinem 1/2. Poniżej zostanie podany wynik analizy dla tego ostatniego przypadku.

5.1.6.1 Wynik dla strumienia cząstek ze spinem 1/2

W Rozdziale 4.3.3.2 pokazaliśmy, że metoda EFI daje dla relatywistycznej cząstki o spinie połówkowym równanie ruchu Diraca (4.52). Pod nieobecność pola elektromagnetycznego wynikające z niego równanie ciągłości ma postać:

$$\frac{\partial}{\partial t} p(\vec{y}, t) + \vec{\nabla} \cdot \vec{P}(\vec{y}, t) = 0, \quad (5.147)$$

gdzie gęstość prawdopodobieństwa p oraz gęstość prądu prawdopodobieństwa są równe odpowiednio:

$$p(\vec{y}, t) = \psi^\dagger \psi, \quad \psi \equiv \psi(\vec{y}, t), \quad (5.148)$$

oraz

$$\vec{P}(\vec{y}, t) = c \psi^\dagger \vec{\alpha} \psi, \quad (5.149)$$

gdzie $\vec{\alpha} \equiv (\alpha^1, \alpha^2, \alpha^3)$ są macierzami Diraca (4.49), natomiast $\psi^\dagger = (\psi_1, \psi_2, \psi_3, \psi_4)^*$ jest polem sprzężonym hermitowsko do bispinora Diraca ψ (Rozdział 4.3.3.2).

Ograniczenie na tempo wzrostu entropii: W [8] pokazano, że tempo wzrostu entropii (5.146) ma w tym przypadku postać:

$$\frac{\partial S_H}{\partial t} \leq c \sqrt{I(t)}. \quad (5.150)$$

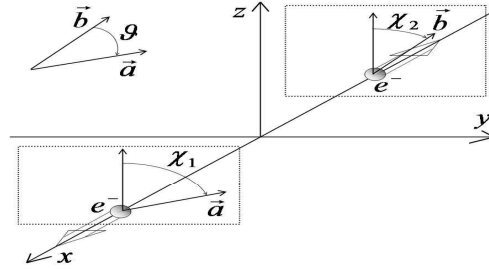
Wniosek: Nierówność ta oznacza, że dla układu, który posiada niezmienniczość przesunięcia, wzrost entropii Shannona rozkładu w jednostce czasu (czyli tempo spadku informacji) jest ograniczony przez skończoną prędkość światła jak również przez pojemność informacyjną, jaką posiada układ (np. swobodny elektron) o swoich współrzędnych czasoprzestrzennych.

Inny sposób interpretacji (5.150) polega na zauważeniu, że dostarcza ona definicji prędkości światła c jako górnego ograniczenia stosunku tempa zmiany entropii do pierwiastka informacji Fishera, która jest przecież na wskroś statystycznym pojęciem informacyjnym.

5.2 Zastosowanie wprowadzonego formalizmu do analizy paradoksu EPR

Przedstawiona w tym rozdziale analiza paradoksu EPR pochodzi od Friedena [8]. Zamieszczamy ją ze względu na to, że efekt ten jest dość powszechnie utożsamiany z własnością teorii kwantowych, ale również z powodu uporządkowania warunków brzegowych zagadnienia zawartego w pracy oryginalnej [8], a przedstawionego w pracy [16] oraz wnioskowania uzgodnionego z przedstawioną w skrypcie fizyczną interpretacją informacji fizycznej K [9, 10].

Opis eksperymentu EPR-Bohm’a: Rozpocznijmy od omówienia eksperymentu EPR-Bohm’a. Rozważmy źródło molekuł o spinie zero, które rozpadają się na parę identycznych cząstek o spinie $1/2$ lecących w przeciwnych kierunkach. Taka, początkowa dla rozważań eksperymentu EPR-Bohm’a, konfiguracja dwucząstkowej molekuly może być efektywnie przygotowana jako stan końcowy w rozpraszaniu $e^-e^- \rightarrow e^-e^-$, gdzie spiny początkowych elektronów procesu są ustawione przeciwnie (równoległe i antyrównoległe względem osi z), a ich początkowe pędy wzdłuż osi y wynoszą \vec{p} oraz $-\vec{p}$ [56]. Istnieje niezerowa amplituda, że dwie rozproszone cząstki (tu wyprodukowane elektrony), poruszają się z pędami wzdłuż osi x , jak na Rysunku 5.1.



Rysunek 5.1: Eksperyment EPR-Bohma. Zaznaczono elektrony rejestrowane w urządzeniach Sterna-Gerlacha; opis w tekście Rozdziału. W lewym górnym rogu rysunku zaznaczono kąt ϑ pomiędzy kierunkami \vec{a} oraz \vec{b} urządzeń Sterna-Gerlacha.

W eksperymencie EPR-Bohm’a dokonywany jest pomiar spinu rozproszonych cząstek, spinu cząstki 1 wzdłuż wektora jednostkowego \vec{a} , tworzącego z osią z kąt χ_1 oraz spinu cząstki 2 wzdłuż wektora jednostkowego \vec{b} , tworzącego z osią z kąt χ_2 . Niech analizator „a”, będący urządzeniem typu Sterna-Gerlacha, mierzy rzut S_a spinu \vec{S}_1 cząstki 1 na kierunek \vec{a} i podobnie analizator „b” mierzy rzut S_b spinu \vec{S}_2 cząstki 2 na kierunek \vec{b} . Kąt pomiędzy płaszczyznami wektorów \vec{a} oraz \vec{b} , zawierającymi oś x , wynosi $\vartheta = \chi_1 - \chi_2$, $0 \leq \vartheta < 2\pi$.

Poniżej wyprowadzimy warunki brzegowe dla metody EFI, uważając aby nie odwoływać się do widzenia rzeczywistości przez pryzmat mechaniki kwantowej⁸.

⁸Lecz biorąc pod uwagę ogólne stwierdzenie, że jeśli P_{12} jest łącznym rozkładem prawdopodobieństwa pewnych dwóch zmiennych losowych „1” oraz „2”, a P_1 oraz P_2 ich rozkładami brzegowymi, to jeśli w ogólności nie są one względem siebie niezależne (tzn. łączne prawdopodobieństwo nie jest iloczynem brzegowych), wtedy:

$$I_F(P_{12}) \geq I_F(P_1) + I_F(P_2) \equiv \tilde{C} \quad (5.151)$$

gdzie \tilde{C} jest pojemnością informacyjną złożonego układu. Relacja (5.151) oznacza, że jeśli występują jakiekolwiek korelacje między zmiennymi to, jeśli znamy wynik doświadczenia dla pierwszej zmiennej to maleje ilość informacji koniecznej do określenia wyniku doświadczenia dla drugiej z nich, tzn. istnienie korelacji w układzie zwiększa informację Fishera I_F o parametrach charakteryzujących rozkład układu.

5.2.1 Warunki brzegowe

W celu rozwiązania równań różniczkowych EFI konieczne jest ustalenie warunków brzegowych na prawdopodobieństwa. Wynikają one z przesłanek fenomenologicznych, zasad zachowania i symetrii przestrzennej badanego układu.

Założenie o istnieniu dokładnie dwóch możliwych rzutów spinu cząstki ze spinem $\hbar/2$ na dowolny wybrany kierunek w przestrzeni: Niech “+” oznacza obserwowaną wartość rzutu spinu $S_a = +\hbar/2$ natomiast “−” oznacza $S_a = -\hbar/2$.

Uwaga o fenomenologii rejestracji spinu cząstki: Jest to jedyne miejsce, w którym uciekamy się do opisu fenomenologicznego, odwołując się do nieklasycznej fizyki zjawiska, mianowicie zakładamy, że rzutowanie spinu cząstki na określony kierunek przestrzenny jest skwantowane zgodnie z wymiarem reprezentacji grupy obrotów (fakt istnienia takich reprezentacji wynika z rozważań entropijnych [8]). Jednak nie oznacza to, że nie istnieje model teoriopowy, który by takie kwantowanie rzutowania opisywał. Musiałby on jedynie zakładać, że po pierwsze spinowe stopnie swobody (z ciągłym rozkładem jego kierunku przed pomiarem) nie mają (prostego) charakteru czasoprzestrzennego [30], a po drugie, że bezwładność cząstki związana ze spinowymi stopniami swobody jest bardzo mała w porównaniu z siłą sprzężenia tych spinowych stopni swobody z otoczeniem (np. aparaturą traktowaną jako rezerwuuar), którego moment pędu ignorujemy. Wtedy przejście cząstki ze spinem przez jakąkolwiek aparaturę pomiarową typu Sterna-Gerlacha porządkowałoby jej spin w sposób dyskretny. Oczywiście zmiana momentu pędu aparatury już nas “nie interesuje”.

Łączną przestrzeń zdarzeń Ω_{ab} stanów spinowych pary cząstek 1 i 2 przyjmujemy jako następującą:

$$S_a S_b \equiv S_{ab} \in \Omega_{ab} = \{S_{++}, S_{--}, S_{+-}, S_{-+}\} \equiv \{(++), (--), (+-), (-+)\} . \quad (5.152)$$

Założmy, że możemy zdefiniować cztery łączne warunkowe prawdopodobieństwa⁹ $P(S_{ab}|\vartheta)$:

$$P(++|\vartheta) , \quad P(--|\vartheta) , \quad P(+-|\vartheta) , \quad P(-+|\vartheta) . \quad (5.153)$$

Warunek normalizacji prawdopodobieństwa $P(S_a S_b|\vartheta)$ w eksperymencie EPR-Bohm’a można, ze względu na wykluczanie się różnych zdarzeń S_{ab} , (5.152), zapisać następująco:

$$P\left(\bigcup_{ab} S_a S_b|\vartheta\right) = \sum_{ab} P(S_a S_b|\vartheta) = 1 \quad \text{dla każdego } \vartheta \in \langle 0, 2\pi \rangle . \quad (5.154)$$

⁹W przypadku braku wspólnej przestrzeni zdarzeń Ω_{AB} dla zmiennych losowych powiedzmy A oraz B , nie można by zdefiniować łącznego rozkładu prawdopodobieństwa $P(A, B)$ dla tych dwóch zmiennych, pomimo istnienia ich rozkładów brzegowych $P(A)$ oraz $P(B)$, co oznacza, że nie dało by się dokonać ich łącznego pomiaru. Również na ogół, pomimo istnienia łącznego rozkładu brzegowego $P(A, B)$ dla zmiennych A oraz B , oraz rozkładu brzegowego $P(B, C)$ dla zmiennych B oraz C , nie istnieje rozkład łączny $P(A, B, C)$ dla zmiennych A , B i C . Zwróćmy uwagę na fakt, że w dowodzie nierówności Bella [57, 48] przyjmuje się za oczywisty fakt istnienia łącznego rozkładu $P(A, B, C)$. Możliwość taka istnieje zawsze, gdy wspólna przestrzeń zdarzeń Ω_{ABC} tych trzech zmiennych losowych istnieje i jest iloczynem kartezjańskim $\Omega_A \times \Omega_B \times \Omega_C$. Z drugiej strony, nierówności Bell’a są znane w klasycznej teorii prawdopodobieństwa od czasów Boole’a jako test, który w przypadku ich niespełnienia świadczy o niemożliwości konstrukcji łącznego rozkładu prawdopodobieństwa. Rozumowanie to można rozszerzyć na dowolną liczbę zmiennych losowych [48, 58].

W pełnym opisie rzeczywistego eksperymentu EPR-Bohm’a, a nie tylko w eksperymencie typu „gedanken”, powinny występować obok dwóch zmiennych losowych rzutów spinów mierzonych w analizatorach „a” oraz „b”, również dwie zmienne losowe kątowe mierzone dla tych cząstek w chwili ich produkcji.

W analizie estymacyjnej postulujemy, że prawdopodobieństwo $P(S_{ab}|\vartheta) \Big|_{\vartheta=\hat{\vartheta}}$ jest funkcją estymatora $\hat{\vartheta}$ parametru ϑ .

Metoda estymacji: Korzystając z proporcjonalności prawdopodobieństw (5.153) do obserwowanej liczności zdarzeń w detektorze, oszacowuje się wartość kąta ϑ . Jednak, aby to uczynić, trzeba mieć model analitycznych formuł na prawdopodobieństwa (5.153). Poniżej wyprowadzimy je metodą EFi.

Sformułowanie warunków brzegowych.

(1) Ponieważ zdarzenia (5.152) wykluczają się wzajemnie i rozpinają całą przestrzeń zdarzeń, więc pierwszym warunkiem brzegowym jest **warunek normalizacji**:

$$\sum_{ab} P(S_{ab}) = P(S_{++}) + P(S_{+-}) + P(S_{-+}) + P(S_{--}) = 1, \quad (5.155)$$

spełniony niezależnie od wartości kąta ϑ .

Dla każdego zdarzenia ab prawdopodobieństwa $P(S_{ab})$ oraz $P(S_{ab}|\vartheta)$ są ze sobą związane następująco:

$$P(S_{ab}) = \int_0^{2\pi} P(S_{ab}|\vartheta) r(\vartheta) d\vartheta, \quad (5.156)$$

gdzie uśrednienie nastąpiło z tzw. *funkcją niewiedzy* $r(\vartheta)$. Ze względu na warunki unormowania (5.154) oraz (5.155) otrzymujemy jej możliwą postać:

$$r(\vartheta) = \frac{1}{2\pi}, \quad 0 \leq \vartheta < 2\pi, \quad (5.157)$$

która oznacza, że w zakresie aparaturowej zmienności ustawienia wartości kąta $\vartheta \in \langle 0, 2\pi \rangle$, z powodu naszej niewiedzy, jest możliwa w równym stopniu każda jego wartość.

Uwaga o wartości kąta ϑ : Wartość ϑ jest związana z ustawieniem aparatury pomiarowej Sterna-Gerlacha „a” i „b”, i trudno ją (na serio) traktować jako wielkość posiadającą rozproszenie. Zasadnicza ϑ jest parametrem charakterystycznym dla przeprowadzonego eksperymentu. Niemniej przez niektórych $r(\vartheta)$ jest widziane jako „prawdopodobieństwo” znane a priori, co oznacza, że wyprowadzoną w ten sposób mechanikę kwantową należałoby traktować jako statystyczną teorię Bayesowską [?].

(2) Kolejne warunki wynikają z **symetrii układu i zasady zachowania całkowitego spinu** przy czym w eksperymencie *względny orbitalny moment pędu wynosi zero*.

Rozważmy prosty przypadek $\vartheta = 0$, gdy obie płaszczyzny, w których ustawione są urządzenia Sterna-Gerlacha są tak samo zorientowane. Z warunku zachowania całkowitego spinu wynika:

$$P(++|0) = P(--|0) = 0, \quad \vartheta = 0 \quad (5.158)$$

co oznacza, że w tym ustawieniu aparatury nigdy nie zobaczymy obu spinów jednocześnie skierowanych w górę czy w dół. Analogicznie, jeśli kąt $\vartheta = \pi$, to warunek:

$$P(+ -|\pi) = P(- +|\pi) = 0, \quad \vartheta = \pi \quad (5.159)$$

oznacza, że w tym przypadku nigdy nie zobaczymy spinów ustawionych jeden w górę drugi w dół. W konsekwencji z zasady zachowania całkowitego spinu otrzymujemy, że jeśli $\vartheta = 0$ lub $\vartheta = \pi$, to zawsze obserwacja jednego spinu daje nam całkowitą wiedzę o drugim. W tym przypadku stany spinów wyraźnie nie są niezależne, są stanami skorelowanymi. Wniosek ten jest intuicyjnie zawarty w sposobie ich przygotowania.

Następnie, ponieważ $P(S_b|\vartheta)$ jest prawdopodobieństwem brzegowym wystąpienia określonej wartości rzutu spinu cząstki 2, zatem nie zależy ono od S_a , czyli od orientacji rzutu spinu cząstki 1, więc nie zależy również od kąta ϑ pomiędzy wektorami \vec{a} oraz \vec{b} . Mamy więc:

$$P(S_b|\vartheta) = C = \text{const.} \quad (5.160)$$

Stąd

$$P(S_b) = \int_0^{2\pi} P(S_b|\vartheta) r(\vartheta) d\vartheta = C \int_0^{2\pi} r(\vartheta) d\vartheta = C \quad \text{dla } S_b = +, - , \quad (5.161)$$

gdzie $r(\vartheta)$ jest określone w (5.157).

Z warunku unormowania prawdopodobieństwa zdarzenia pewnego, mamy

$$P(S_b = +) + P(S_b = -) = C + C = 2C = 1 \quad \text{czyli } C = \frac{1}{2} . \quad (5.162)$$

Zatem z (5.160) otrzymujemy, że¹⁰ :

$$P(S_b|\vartheta) = \frac{1}{2} , \quad (5.164)$$

natomiast z (5.161):

$$P(S_b) = \frac{1}{2} . \quad (5.165)$$

Inną ważną własnością symetrii przestrzennej eksperymentu jest brak preferencji występowania spinu skierowanego w górę czy w dół, tzn:

$$P(S_{+-}|\vartheta) = P(S_{-+}|\vartheta) \quad \text{oraz} \quad P(S_{++}|\vartheta) = P(S_{--}|\vartheta) , \quad (5.166)$$

co oznacza, że jeśli obserwowalibyśmy eksperyment dla układu odwróconego względem osi x o kąt π , to statystyczny wynik byłby dokładnie taki sam. Z (5.166) otrzymujemy:

$$P(S_{++}) = \int P(S_{++}|\vartheta) r(\vartheta) d\vartheta = \int P(S_{--}|\vartheta) r(\vartheta) d\vartheta = P(S_{--}) \quad (5.167)$$

oraz w sposób analogiczny:

$$P(S_{+-}) = P(S_{-+}) . \quad (5.168)$$

¹⁰Podobnie dla S_a , wychodząc z $P(S_a|\vartheta) = C' = \text{const.}$ i postępując analogicznie jak przy przejściu od (5.160) do (5.162), otrzymujemy:

$$P(S_a|\vartheta) = \frac{1}{2} . \quad (5.163)$$

Dodatkowo otrzymujemy:

$$\begin{aligned} P(S_{+-}) &= \int P(S_{+-}|\vartheta) r(\vartheta) d\vartheta = \int P(S_{+-}|\vartheta + \pi) r(\vartheta + \pi) d\vartheta \\ &= \int P(S_{++}|\vartheta) r(\vartheta + \pi) d\vartheta = P(S_{++}) , \end{aligned} \quad (5.169)$$

gdzie w drugiej równości skorzystano ze zwykłej zamiany zmiennych $\vartheta \rightarrow \vartheta + \pi$, w trzeciej równości z tego, że $P(S_{+-}|\vartheta + \pi) = P(S_{++}|\vartheta)$, a w ostatniej z $r(\vartheta) = 1/(2\pi) = r(\vartheta + \pi)$.

Jako konsekwencja (5.167)-(5.169) otrzymujemy:

$$P(S_{-+}) = P(S_{--}) . \quad (5.170)$$

Ze względu na (5.167)-(5.170) oraz (5.155) otrzymujemy:

$$P(S_{ab}) = \frac{1}{4} . \quad (5.171)$$

W końcu wzór Bayes'a na *prawdopodobieństwo warunkowe* daje:

$$P(S_a|S_b) = \frac{P(S_{ab})}{P(S_b)} = \frac{1/4}{1/2} = \frac{1}{2} \quad \text{dla każdego } S_a, S_b . \quad (5.172)$$

Końcowe uwagi o warunkach brzegowych: Zauważmy, że jak dotąd w wyprowadzeniu zależności (5.158) do (5.172) nie skorzystano z EFL. Są to bowiem warunki brzegowe dla równań metody EFL. Zależności te wynikały z początkowej obserwacji istnienia dla cząstki o spinie połówkowym dokładnie dwóch możliwych rzutów spinu na dowolny kierunek (por. Uwaga na początku rozdziału), zasady zachowania całkowitego momentu pędu oraz z symetrii układu. Przy wyborze rozwiązania równania generującego rozkład, skorzystamy jeszcze z warunku geometrycznej symetrii układu przy obrocie o kąt $\vartheta = 2\pi$.

5.2.2 Pojemność informacyjna dla zagadnienia EPR-Bohm'a

Poniżej wyprowadzimy wyrażenia dla prawdopodobieństw (5.153) jako wynik estymacji rozkładów metodą EFL.

Wielkość mierzona przez obserwatora: Obserwator zewnętrzny mierzy jedynie wartość rzutu spinu jednej cząstki *powiedzmy*, że cząstki 1, to znaczy S_a , natomiast nie mierzy wartości rzutu spinu drugiej cząstki S_b , jest ona traktowana jako wielkość nieznana, ale ustalona. Również wartość S_b nie jest estymowana z obserwacji S_a .

Określenie przestrzeni położeń dla EFL: Podobnie jak poprzednio układ sam próbkuje swoimi Fisherowskimi kinetycznymi stopniami swobody dostępną mu przestrzeń położeń. Tym razem jest to jednowymiarowa przestrzeń kąta ϑ .

Na podstawie danych pomiarowych wartości rzutu spinu S_a , estymujemy metodą EFL kąt ϑ pomiędzy dwoma płaszczyznami zaznaczonymi na rysunku 5.1. Kąt ten jest więc traktowany jako nieznaną parametr i jest estymowany z obserwacji S_a .

Określenie funkcji wiarygodności i przestrzeni próby: Ponieważ dokonujemy pomiaru S_a , a S_b oraz ϑ są nieznanymi, ale ustalonymi wielkościami, zatem funkcję wiarygodności dla powyżej postawionego problemu zapiszemy następująco:

$$P(S_a|S_b, \vartheta) \rightarrow \text{funkcja wiarygodności próby dla } \vartheta. \quad (5.173)$$

Jej postaci szukamy, wykorzystując metodę EFL. Jednak postać informacji Fishera, która byłaby miarą precyzji estymacji kąta ϑ opartej o pomiar S_a , wynikać będzie z jej pierwotnej postaci (2.193) dla wymiaru próby $N = 1$.

Skoro więc ϑ jest estymowanym parametrem, zatem pełni on teraz taką rolę jak parametrem θ w wzorze (2.193), w którym, obok podstawienia $\theta \rightarrow \vartheta$, należy dokonać następujących podstawień:

$$n = N = 1 \quad \text{oraz} \quad \int dy^1 \rightarrow \sum_a, \quad (5.174)$$

gdzie przestrzenią (jednowymiarowej) próby jest zbiór $\{-1, +1\}$ wartości zmiennej losowej S_a .

Amplitudy prawdopodobieństwa q_{ab} są zdefiniowane jak zwykle następująco:

$$P(S_a|S_b, \vartheta) = q_{ab}^2(\vartheta). \quad (5.175)$$

Oczekiwana IF parametru ϑ . Pojemność informacyjna kanału (ϑ, S_b) jest równa:

$$\begin{aligned} I_b &= \sum_{a=-}^{+} \frac{1}{P(S_a|S_b, \vartheta)} \left(\frac{\partial P(S_a|S_b, \vartheta)}{\partial \vartheta} \right)^2 = \sum_{a=-}^{+} \frac{1}{q_{ab}^2} \left(\frac{\partial q_{ab}^2}{\partial \vartheta} \right)^2 = \sum_{a=-}^{+} \frac{1}{q_{ab}^2} \left(2q_{ab} \frac{\partial q_{ab}}{\partial \vartheta} \right)^2 \\ &= 4 \sum_{a=-}^{+} \left(\frac{\partial q_{ab}}{\partial \vartheta} \right)^2 \end{aligned} \quad (5.176)$$

i podsumowując:

$$I_b \equiv I(S_b, \vartheta) = 4 \sum_{a=-}^{+} q_{ab}'^2(\vartheta), \quad S_b = (+, -), \quad \text{gdzie} \quad q_{ab}' \equiv \frac{dq_{ab}}{d\vartheta}, \quad (5.177)$$

gdzie sumowanie po a odpowiada całkowaniu w informacji Fishera po przestrzeni bazowej.

Zwróćmy uwagę, że ϑ w powyższej formule na $I_b \equiv I(S_b, \vartheta)$ jest wyznaczone dla konkretnej wartości ϑ i konkretnej wartości S_b rzutu spinu cząstki 2. Zatem jest to pojemność informacyjna jednego kanału (ϑ, S_b) czyli *informacja Fishera parametru ϑ w tym kanale*:

$$I_F(\vartheta) = I_b. \quad (5.178)$$

Powróćmy do warunków (5.174). Pierwszy z nich oznacza, że ranga amplitudy pola jest równa $N = 1$, drugi oznacza, że *przestrzeń bazowa to teraz dwupunktowy zbiór wartości rzutu spinu cząstki 1 na kierunek 'a'*. W poniższych uwagach odniesiemy się do tych elementów analizy.

Uwaga o sumowaniu po a w przestrzeni próby: Jako, że ϑ jest parametrem, a pomiary w próbie są związane z obserwacjami S_a , zatem zapis w (5.177) wymaga pewnego wyjaśnienia. Wyrażenie

(5.177) reprezentuje dwa równania dla dwóch możliwych wartości S_b . Ponieważ zmienną w pomiarze jest rzut spinu cząstki 1, zatem sumowanie przebiega po dwóch możliwych wartościach spinu $S_a = +, -$. Tak więc, zapis $I(S_b, \vartheta)$ wskazuje na *informację zawartą w przestrzeni próby zmiennej S_a dla cząstki 1 na temat nieznanego kąta ϑ w obecności pewnej nieznannej, lecz ustalonej wartości rzutu spinu S_b cząstki 2*.

Całkowita pojemność informacyjna I_{1a} dla parametru ϑ : Ponieważ kąt ϑ jest parametrem, którego wartość może się zmieniać w sposób ciągły w przedziale $\langle 0, 2\pi \rangle$, stąd zgodnie z (5.177) mamy nieskończoną liczbę kanałów informacji Fishera związanych z wartościami ϑ . Dla każdego z tej nieskończonej liczby kanałów są jeszcze dwa kanały informacyjne związane z możliwymi wartościami dla S_b . Aby poradzić sobie z taką sytuacją metoda EFI wykorzystuje pojedynczą, skalarną informację (oznaczymy ją I_{1a}), nazywaną pojemnością informacyjną, wprowadzoną w Rozdziale 2.7. Wielkość tą konstruujemy dokonując sumowania informacji po wszystkich możliwych kanałach. Zatem, suma przebiegać będzie po wszystkich wartościach kąta ϑ_k przy nieznanym rzucie spinu S_b cząstki 2. Tak więc, po pierwsze, pojemność informacyjna I_{nk} dla jednego kanału, wprowadzona po raz pierwszy w (2.207), przyjmuje postać:

$$I_{bk} = I(S_b, \vartheta_k) . \quad (5.179)$$

Po drugie, w wyniku sumowania po wszystkich możliwych wartościach kąta ϑ_k oraz rzutach spinu S_b cząstki 2, otrzymujemy *całkowitą pojemność informacyjną dla parametru ϑ* :

$$\begin{aligned} I_{1a} &\equiv \sum_{b=-}^{+} \sum_k I(S_b, \vartheta_k) \rightarrow \\ &\rightarrow I_{1a} = \sum_{b=-}^{+} \int_0^{2\pi} d\vartheta I(S_b, \vartheta) = 4 \sum_{b=-}^{+} \sum_{a=-}^{+} \int_0^{2\pi} d\vartheta q_{ab}'^2(\vartheta) \equiv 4 \sum_{ab} \int_0^{2\pi} d\vartheta q_{ab}'^2(\vartheta) , \end{aligned} \quad (5.180)$$

gdzie całkowanie pojawia się z powodu zastąpienia sumy dla dyskretnego indeksu k całkowaniem po ciągłym zbiorze wartości parametru ϑ . Oznacza to, że po zrozumieniu czym jest pojedynczy k - ty kanał związany z ϑ , wykonujemy w celu otrzymania całkowitej pojemności informacyjnej całkowanie, które przebiega zgodnie z (5.157) od 0 do 2π . W ostatniej linii wykorzystano (5.177).

Sumowanie w (5.180) po ab przebiega po wszystkich możliwych kombinacjach łącznej przestrzeni stanów S_{ab} określonej w (5.152). Indeks '1_a' w I_{1a} oznacza, że jest to pojemność informacyjna jednej cząstki, przy czym wyróżniliśmy cząstkę 1, dla której pomiar dokonywany jest w analizatorze a . **Pojemność informacyjna I_{1a} wchodzi do estymacyjnej procedury EFI.**

Uwaga o randze $N = 1$. Zachodzi pytanie: Jeśli okaże się, że tradycyjne formuły mechaniki kwantowej dla eksperymentu EPR-Bohm'a pojawiają się dla powyżej określonej estymacji z $N = 1$ (a tak się istotnie stanie), to czyżby pomiar oddziaływania cząstki z aparaturą Sterna-Gerlach'a miał mieć termodynamiczny charakter oddziaływania małego układu z termostatem, w rozumieniu podanym na samym początku Rozdziału 5.2.1? W końcu, jak wiemy z poprzednich rozdziałów, rozwiązania z $N = 1$ odnoszą się do zjawisk termodynamicznych.

5.2.3 Informacja strukturalna. Amplituda prawdopodobieństwa

Biorąc pod uwagę ogólną postać informacji strukturalnej (3.25) oraz uwagi Rozdziału 5.2.2 odnośnie konstrukcji postaci I_{1_a} , (5.180), zauważamy, że *informacja strukturalna* Q_{1_a} dla obserwowanej cząstki ma postać:

$$Q_{1_a} \equiv \sum_{ab} \int_0^{2\pi} d\vartheta q_{ab}^2(\vartheta) \Phi_{ab}(q_{ab}) . \quad (5.181)$$

Poniższa analiza związana z wyprowadzeniem amplitud q_{ab} jest podobna do analizy dla rozkładu Boltzmann w Rozdziale 5.1.

Wariacyjna zasada informacyjna: Dla pojemności informacyjnej I , (5.180), oraz informacji strukturalnej Q , (5.181) wariacyjna zasada informacyjna ma postać:

$$\delta_{(q_{ab})} K \equiv \delta_{(q_{ab})} (I_{1_a} + Q_{1_a}) = \delta_{(q_{ab})} \left(\int_0^{2\pi} d\vartheta k \right) = 0 , \quad (5.182)$$

gdzie zgodnie z ogólną postacią (4.7), gęstość całkowitej informacji fizycznej k dla amplitud q_{ab} jest równa:

$$k = 4 \sum_{ab} \left(q_{ab}'^2 + \frac{1}{4} q_{ab}^2(\vartheta) \Phi_{ab}(q_{ab}) \right) . \quad (5.183)$$

Rozwiązaniem problemu *wariacyjnego* (5.182) względem q_{ab} jest *równanie Eulera-Lagrange'a* (4.12) (por. (5.16)):

$$\frac{d}{d\vartheta} \left(\frac{\partial k}{\partial q_{ab}'(\vartheta)} \right) = \frac{\partial k}{\partial q_{ab}} . \quad (5.184)$$

Z równania (5.184) dla k jak w (5.183) otrzymujemy, dla każdego dwucząstkowego łącznego stanu spinowego S_{ab} :

$$q_{ab}'' = \frac{1}{2} \frac{d(\frac{1}{4} q_{ab}^2 \Phi_{ab}(q_{ab}))}{dq_{ab}} . \quad (5.185)$$

Ponieważ $q_{ab}^2(\vartheta) \Phi_{ab}(q_{ab})$ jest jawnie jedynie funkcją q_{ab} , więc różniczka zupełna zastąpiła pochodną cząstkową po q_{ab} występującą w (5.184).

Zmodyfikowana obserwowana zasada strukturalna: Po wycałkowaniu (5.180) przez części, pojemność I wynosi (por. (5.18), (5.20), (5.22)):

$$I_{1_a} = 4 \sum_{ab} \int_0^{2\pi} d\vartheta \left(\tilde{C}_{ab} - q_{ab} q_{ab}'' \right) , \quad (5.186)$$

gdzie

$$\tilde{C}_{ab} = \frac{1}{2\pi} \left(q_{ab}(2\pi) q_{ab}'(2\pi) - q_{ab}(0) q_{ab}'(0) \right) . \quad (5.187)$$

Z powodu braku dodatkowych więzów $\kappa = 1$, zatem zmodyfikowana obserwowana zasada strukturalna $\tilde{i}' + \tilde{\mathbf{C}}_{EPR} + q = 0$, (3.43), jest ze względu na (5.186) oraz (5.181), następująca:

$$4 \sum_{ab} \left(-q_{ab} q''_{ab} + \tilde{C}_{ab} + \frac{1}{4} q_{ab}^2 \mathbf{F}_{ab}(q_{ab}) \right) = 0. \quad (5.188)$$

Poniżej przekonamy się, że $\tilde{\mathbf{C}}_{EPR} = 4 \sum \tilde{C}_{ab} = 0$.

Dla każdego dwucząstkowego łącznego stanu spinowego S_{ab} , obserwowana zasada strukturalna (5.188) ma więc postać (por. 5.24):

$$-q_{ab} q''_{ab} + \tilde{C}_{ab} + \frac{1}{4} q_{ab}^2 \mathbf{F}_{ab}(q_{ab}) = 0. \quad (5.189)$$

Wraz z równaniem Eulera-Lagrange'a (5.185), równanie (5.189) posłuży do wyprowadzenia równania generującego rozkład.

Wyprowadzenia równania generującego: Wykorzystując w (5.189) związek (5.185), otrzymujemy (por. (5.25)) dla każdego dwucząstkowego łącznego stanu spinowego S_{ab} :

$$\frac{1}{2} q_{ab} \frac{d(\frac{1}{4} q_{ab}^2 \mathbf{F}_{ab}(q_{ab}))}{dq_{ab}} = \frac{1}{4} q_{ab}^2 \mathbf{F}_{ab}(q_{ab}) + \tilde{C}_{ab}. \quad (5.190)$$

Zapiszmy powyższe równanie w wygodniejszej formie:

$$\frac{2dq_{ab}}{q_{ab}} = \frac{d(\frac{1}{4} q_{ab}^2 \mathbf{F}_{ab}(q_{ab}))}{\frac{1}{4} q_{ab}^2 \mathbf{F}_{ab}(q_{ab}) + \tilde{C}_{ab}}, \quad (5.191)$$

z której po obu stronach wyciągnięciu otrzymujemy:

$$\frac{1}{4} q_{ab}^2(\vartheta) \mathbf{F}_{ab}(q_{ab}) = \frac{q_{ab}^2(\vartheta)}{A_{ab}^2} - \tilde{C}_{ab}, \quad (5.192)$$

gdzie A_{ab}^2 jest w ogólności zespoloną stałą.

Równanie generujące: Podstawiając (5.192) do (5.185) otrzymujemy szukane różniczkowe równanie generujące dla amplitud q_{ab} :

$$q''_{ab}(\vartheta) = \frac{q_{ab}(\vartheta)}{A_{ab}^2}, \quad (5.193)$$

będące konsekwencją obu zasad informacyjnych, strukturalnej i wariacyjnej.

Rozwiązanie równania generującego: Ponieważ amplituda q_{ab} jest rzeczywista, więc A_{ab}^2 też musi być rzeczywiste. Podobnie jak w rozdziale 5.1.3, ponieważ stała A_{ab}^2 jest rzeczywista, więc można ją przedstawić za pomocą innej rzeczywistej stałej a_{ab} jako $A_{ab} = a_{ab}$ lub $A_{ab} = i a_{ab}$. Zatem istnieją dwie klasy rozwiązań równania (5.193).

Dla $A_{ab} = a_{ab}$, rozwiązanie (5.193) ma charakter czysto eksponencjalny:

$$q_{ab}(\vartheta) = B''_{ab} \exp\left(-\frac{\vartheta}{a_{ab}}\right) + C''_{ab} \exp\left(\frac{\vartheta}{a_{ab}}\right), \quad A_{ab} = a_{ab}, \quad (5.194)$$

gdzie stałe B''_{ab} oraz C''_{ab} są rzeczywiste.

Natomiast dla $A_{ab} = i a_{ab}$, rozwiązanie (5.193) ma charakter *trygonometryczny*:

$$q_{ab}(\vartheta) = B'_{ab} \sin\left(\frac{\vartheta}{a_{ab}}\right) + C'_{ab} \cos\left(\frac{\vartheta}{a_{ab}}\right), \quad A_{ab} = i a_{ab}, \quad (5.195)$$

gdzie a_{ab} , B'_{ab} , C'_{ab} są rzeczywistymi stałymi.

Warunek niezmienniczości przy obrocie o 2π : W rozważanym obecnie przypadku¹¹ wartości ϑ są kątem z ograniczonego zbioru $\langle 0, 2\pi \rangle$. Zatem z geometrycznej symetrii układu przy obrocie o kąt 2π wynika, że rozkład $P(S_a|S_b, \vartheta)$ jest również funkcją okresową zmiennej ϑ . Ze względu na (5.175) warunek ten oznacza, że funkcja $q_{ab}(\vartheta)$ powinna być okresowa, *zatem wybieramy rozwiązanie o charakterze trygonometrycznym*, przy czym Funkcje sin oraz cos w (5.195) są funkcjami bazowymi tworzącymi amplitudę prawdopodobieństwa $q_{ab}(\vartheta)$.

Funkcje bazowe na przestrzeni amplitud powinny być ortogonalne. Z postaci q_a w (5.195) wynika, że funkcjami bazowymi są funkcje 'sin' oraz 'cos'. Ponieważ q_a jest określone na przestrzeni parametru $\vartheta \in \langle 0, 2\pi \rangle$, zatem *warunek ortogonalności funkcji bazowych* na tej przestrzeni:

$$\int_0^{2\pi} d\vartheta \sin(\vartheta/a_{ab}) \cos(\vartheta/a_{ab}) = 0 \quad (5.196)$$

daje po wycalkowaniu postać stałych a_{ab} :

$$a_{ab} = \frac{2}{n_{ab}}, \quad \text{gdzie } n_{ab} = 1, 2, \dots \quad (5.197)$$

Warunek minimalizacji pojemności I : Warunek (5.197) jest warunkiem na dopuszczalne wartości $a_{ab} = -iA_{ab}$ (por. (5.195)), ale jak widać nie wskazuje on na żadną z nich jednoznacznie. Może to nastąpić, gdy ustalimy wartość n_{ab} , co uczynimy ograniczając rozważania do zapostulowanego w Rozdziale 2.7.1 warunku minimalizacji informacji kinetycznej $I \rightarrow \min$.

Zacznijmy od wyznaczenia stałej \tilde{C}_{ab} , (5.187). Po wstawieniu do (5.187) amplitudy (5.195) z (5.197), otrzymujemy:

$$\begin{aligned} \tilde{C}_{ab} &= \frac{n_{ab}}{4\pi} \sin(\pi n_{ab}) \\ &\cdot \left((B'_{ab} - C'_{ab})(B'_{ab} + C'_{ab}) \cos(\pi n_{ab}) - 2B'_{ab}C'_{ab} \sin(\pi n_{ab}) \right) = 0, \quad \text{gdzie } n_{ab} = 1, 2, \dots, \end{aligned} \quad (5.198)$$

skąd po skorzystaniu z $I_{1a} = 4 \sum \int d\vartheta (\tilde{C}_{ab} - q_{ab} q''_{ab})$, (5.186), oraz równania generującego (5.193) pozwalającego wyeliminować q''_{ab} , otrzymujemy użyteczną postać pojemności informacyjnej:

$$I_{1a} = -4 \sum_{ab} \frac{1}{A_{ab}^2} \int_0^{2\pi} d\vartheta q_{ab}^2(\vartheta). \quad (5.199)$$

¹¹Przypomnijmy sobie analizę przeprowadzoną w rozdziale (5.1.3) dla rozkładu energii cząsteczki gazu. Wtedy ze względu na nieograniczony zakres argumentu amplitudy, wybraliśmy rozwiązanie o charakterze eksponencjalnym.

Należy jeszcze wyznaczyć całkę w (5.199):

$$\begin{aligned}
\int_0^{2\pi} d\vartheta q_{ab}^2(\vartheta) &\equiv \int_0^{2\pi} d\vartheta P(S_a|S_b, \vartheta) = \int_0^{2\pi} d\vartheta \frac{P(S_a, S_b, \vartheta)}{p(\vartheta, S_b)} = \int_0^{2\pi} d\vartheta \frac{p(\vartheta, S_a|S_b) P(S_b)}{p(\vartheta, S_b)} \\
&= \int_0^{2\pi} d\vartheta \frac{p(\vartheta, S_a|S_b)^{\frac{1}{2}}}{p(S_b|\vartheta) r(\vartheta)^{\frac{1}{2}}} = \int_0^{2\pi} d\vartheta \frac{p(\vartheta, S_a|S_b)^{\frac{1}{2}}}{\frac{1}{2} \frac{1}{2\pi}} \\
&= 2\pi \int_0^{2\pi} d\vartheta p(\vartheta, S_a|S_b) = 2\pi P(S_a|S_b) = 2\pi \frac{1}{2} = \pi,
\end{aligned} \tag{5.200}$$

gdzie w pierwszej linii i) wpięrow skorzystano z definicji amplitudy, $q_{ab}^2(\vartheta) = P(S_a|S_b, \vartheta)$, (5.175), ii) następnie z definicji prawdopodobieństwa warunkowego, iii) znowu z definicji prawdopodobieństwa warunkowego:

$$p(\vartheta, S_a|S_b) = \frac{P(S_a, S_b, \vartheta)}{P(S_b)}, \tag{5.201}$$

następnie w drugiej linii z $p(\vartheta, S_b) = p(S_b|\vartheta) r(\vartheta)$ (por. Uwaga poniżej (5.157)) oraz z wyrażen (5.157), (5.164) oraz w trzeciej linii z:

$$P(S_a|S_b) = \int_0^{2\pi} d\vartheta p(\vartheta, S_a|S_b), \tag{5.202}$$

a na koniec z (5.172).

Wykorzystując (5.199), (5.200) oraz $A_{ab} = i a_{ab}$ i (5.197), możemy wyrazić informację I_{1_a} poprzez stałe n_{ab} , otrzymując:

$$I_{1_a} = -Q_{1_a} = -4 \sum_{ab} \frac{1}{A_{ab}^2} \int_0^{2\pi} d\vartheta q_{ab}^2 = 4\pi \sum_{ab} \frac{1}{a_{ab}^2} = \pi \sum_{ab} n_{ab}^2, \text{ gdzie } n_{ab} = 1, 2, \dots, \tag{5.203}$$

przy czym drugą z powyższych równości, a zatem i pierwszą, otrzymano korzystając z postaci informacji strukturalnej (5.181), (5.192) oraz z (5.198). Związek $Q_{1_a} = -I_{1_a}$ jest wyrazem ogólnego warunku (3.45) spełnienia przez metodę EFI oczekiwanej zasady strukturalnej (3.35).

Z powyższego związku wynika, że warunek minimalizacji I_{1_a} będzie spełniony dla:

$$n_{ab} = 1, \quad I_{1_a} \rightarrow \min, \tag{5.204}$$

dla dowolnych S_a oraz S_b . Warunek $n_{ab} = 1$ zgodnie z (5.197) odpowiada następującej wartości a_{ab} :

$$a_{ab} = 2 \quad \text{dla dowolnych } S_a \text{ oraz } S_b. \tag{5.205}$$

Sumowanie w (5.203) przebiega po wszystkich S_a i S_b , zatem dla $n_{ab} = 1$, otrzymujemy wartość minimalną I_{1_a} :

$$I_{1_a}(\min) = \pi \sum_{a=-}^{+} \sum_{b=-}^{+} n_{ab}^2 = 4\pi, \text{ gdzie } n_{ab} = 1. \tag{5.206}$$

Tak więc otrzymaliśmy minimalną wartość pojemności informacyjnej dla parametru ϑ .

Wyznaczenie stałych w amplitudzie: W wyrażeniu (5.195), które dla $a_{ab} = 2$ przyjmuje postać:

$$q_{ab}(\vartheta) = B'_{ab} \sin\left(\frac{\vartheta}{2}\right) + C'_{ab} \cos\left(\frac{\vartheta}{2}\right), \quad (5.207)$$

występują jeszcze stałe B'_{ab} oraz C'_{ab} , które również musimy wyznaczyć. Wyznaczymy je korzystając z wcześniej ustalonych w (5.158), wartości łącznego prawdopodobieństwa warunkowego $P(S_{ab}|\vartheta)$ dla $\vartheta = 0$.

Wyrażmy wpierw $P(S_{ab}|\vartheta)$ poprzez amplitudę q_{ab} :

$$\begin{aligned} P(S_{ab}|\vartheta) &\equiv P(S_a S_b|\vartheta) = \frac{P(S_a S_b, \vartheta)}{r(\vartheta)} = P(S_a|S_b, \vartheta) \frac{P(S_b, \vartheta)}{r(\vartheta)} \\ &= P(S_a|S_b, \vartheta) P(S_b|\vartheta) = q_{ab}^2(\vartheta) P(S_b|\vartheta) = \frac{1}{2} q_{ab}^2(\vartheta), \end{aligned} \quad (5.208)$$

gdzie skorzystano z (5.164). Z równania (5.208) wynika, że:

$$q_{ab}(\vartheta) = 0 \quad \text{jeśli} \quad P(S_{ab}|\vartheta) = 0. \quad (5.209)$$

Korzystając z (5.209) oraz $P(++|0) = P(--|0) = 0$, (5.158), i wstawiając (5.207) dla $\vartheta = 0$ do (5.208), otrzymujemy:

$$C'_{++} = C'_{--} = 0. \quad (5.210)$$

Ponadto, korzystając z symetrii geometrycznej eksperymentu, $P(S_{+-}|\vartheta) = P(S_{-+}|\vartheta)$ oraz $P(S_{++}|\vartheta) = P(S_{--}|\vartheta)$, zapisanej w (5.166), otrzymujemy:

$$B'_{++} = B'_{--} \quad \text{oraz} \quad B'_{+-} = B'_{-+}, \quad (5.211)$$

oraz

$$C'_{+-} = C'_{-+}. \quad (5.212)$$

Analiza z wykorzystaniem metryki Rao-Fishera: Poniżej przekonamy się, że wyznaczenie pozostałych stałych B'_{ab} oraz C'_{ab} wymaga dodatkowego założenia, odnoszącego się do postaci metryki Rao-Fishera na przestrzeni statystycznej \mathcal{S} , szukanego statystycznego modelu.

Szukanym rozkładem prawdopodobieństwa w eksperymencie EPR-Bohm'a jest dyskretny rozkład (5.153) określony na przestrzeni zdarzeń $S_a S_b \equiv S_{ab} \in \Omega_{ab}$, (5.152), i unormowany zgodnie z (5.154) do jedności. Zatem zbiór możliwych wyników to:

$$i \equiv ab = (++), (--), (+-), (-+) . \quad (5.213)$$

Aplitudy prawdopodobieństwa związane z rozkładem (5.153) mają zgodnie z (5.208) postać:

$$\tilde{q}_i \equiv \frac{1}{\sqrt{2}} q_{ab}(\vartheta) = \sqrt{P(S_{ab}|\vartheta)}. \quad (5.214)$$

Zgodnie z (5.207) amplitudy $q_{ab}(\vartheta)$ mają następujące pochodne:

$$\frac{\partial q_{ab}}{\partial \vartheta} = \frac{1}{2} \left(B'_{ab} \cos\left(\frac{\vartheta}{2}\right) - C'_{ab} \sin\left(\frac{\vartheta}{2}\right) \right). \quad (5.215)$$

Dla $\aleph = 4$ wyników (5.213), z ogólnego związku (2.173) określającego metrykę Rao-Fishera $g_{ab} = 4 \sum_{i=1}^{\aleph} \frac{\partial q^i}{\partial \theta^a} \frac{\partial q^i}{\partial \theta^b}$, otrzymujemy po skorzystaniu z (5.214) oraz (5.215) następującą postać metryki $g_{\vartheta\vartheta}$ indukowanej z rozkładu (5.153):

$$\begin{aligned} g_{\vartheta\vartheta}(\vartheta) &= 4 \sum_{i=1}^{\aleph=4} \frac{\partial \tilde{q}_i}{\partial \vartheta} \frac{\partial \tilde{q}_i}{\partial \vartheta} = 4 \sum_{ab} \frac{\partial(\frac{1}{\sqrt{2}}q_{ab})}{\partial \vartheta} \frac{\partial(\frac{1}{\sqrt{2}}q_{ab})}{\partial \vartheta} \\ &= \frac{1}{2} \sum_{ab} \left((B'_{ab})^2 + ((C'_{ab})^2 - (B'_{ab})^2) \sin^2(\frac{\vartheta}{2}) - B'_{ab} C'_{ab} \sin(\vartheta) \right) \end{aligned} \quad (5.216)$$

na jednowymiarowej przestrzeni statystycznej \mathcal{S} parametryzowanej w bazie ϑ .

Po skorzystaniu z (5.211), postać metryki $g_{\vartheta\vartheta}(\vartheta)$, (5.216), prowadzi dla $\vartheta = 0$ do warunku:

$$g_{\vartheta\vartheta}(\vartheta = 0) = \frac{1}{2} \sum_{ab} (B'_{ab})^2 = (B'_{++})^2 + (B'_{+-})^2 \quad \text{dla } \vartheta = 0, \quad (5.217)$$

natomiast dla $\vartheta = \pi$ do warunku:

$$g_{\vartheta\vartheta}(\vartheta = \pi) = \frac{1}{2} \sum_{ab} (C'_{ab})^2 = (C'_{+-})^2 \quad \text{dla } \vartheta = \pi, \quad (5.218)$$

gdzie skorzystano również z (5.210).

Centralne założenie statystyczne dla eksperymentu EPR-Bohm'a: Załóżmy, że w eksperymencie EPR-Bohm'a metryka Rao-Fishera $g_{\vartheta\vartheta}$ na \mathcal{S} jest niezależna od wartości parametru ϑ , tzn.:

$$g_{\vartheta\vartheta}(\vartheta) = g_{\vartheta\vartheta} = \text{const.} \quad (5.219)$$

W szczególnym przypadku warunek (5.219) oznacza, że $g_{\vartheta\vartheta}(\vartheta = 0) = g_{\vartheta\vartheta}(\vartheta = \pi)$, co uwzględniając w zależnościach (5.217) oraz (5.218) daje:

$$(B'_{++})^2 + (B'_{+-})^2 = (C'_{+-})^2 \geq 0. \quad (5.220)$$

Warunek ten oznacza, że $C'_{+-} \neq 0$, gdyż w przeciwnym wypadku, tzn. dla $C'_{+-} = 0$, z warunku (5.220) oraz z (5.210) otrzymalibyśmy zerowanie się wszystkich współczynników B'_{ab} oraz C'_{ab} , co odpowiadałoby trywialnemu przypadkowi braku rozwiązania dla zagadnienia EPR. Zatem otrzymujemy:

$$C \equiv C'_{+-} = C'_{-+} \neq 0, \quad (5.221)$$

gdzie równości współczynników wynika z (5.212).

W końcu niezależność $g_{\vartheta\vartheta}$ od wartości ϑ , (5.219), daje po skorzystaniu z postaci $g_{\vartheta\vartheta}$, (5.216), oraz z warunków (5.210), (5.211) i (5.212), warunek:

$$\sum_{ab} \left((C'_{ab})^2 - (B'_{ab})^2 \right) = 2 \left((C'_{+-})^2 - (B'_{++})^2 - (B'_{+-})^2 \right) = 0, \quad (5.222)$$

czyli warunek pokrywający się z (5.220) oraz:

$$\sum_{ab} B'_{ab} C'_{ab} = 2(B'_{+-} C'_{+-}) = 0. \quad (5.223)$$

Warunek (5.223) wraz z (5.221) oznacza:

$$B'_{+-} = B'_{-+} = 0, \quad (5.224)$$

gdzie skorzystano również z (5.211). Po uwzględnieniu (5.224) w (5.222), otrzymujemy:

$$(B'_{++})^2 = (C'_{+-})^2. \quad (5.225)$$

W końcu zauważmy, że ze względu na (5.224) oraz (5.221), warunek istnienia nietrywialnego rozwiązania oznacza, że:

$$B \equiv B'_{++} = B'_{--} \neq 0, \quad (5.226)$$

gdzie ponownie w równości skorzystano z (5.211).

Podstawiając otrzymane wyniki dla współczynników B'_{ab} oraz C'_{ab} do (5.207), otrzymujemy:

$$q_{++}(\vartheta) = q_{--}(\vartheta) = B \sin(\vartheta/2), \quad q_{-+}(\vartheta) = q_{+-}(\vartheta) = C \cos(\vartheta/2). \quad (5.227)$$

Jak widać, równość współczynników w związkach (5.221) oraz (5.226) jest odbiciem równości odpowiednich amplitud, wynikającej z symetrii odbicia przestrzennego (5.166) oraz wzoru (5.208).

Musimy jeszcze wyznaczyć stałe B oraz C . Z powodu warunku normalizacji prawdopodobieństwa $P(S_a S_b | \vartheta)$, (5.154), otrzymujemy ze względu na (5.208) równanie:

$$\frac{1}{2} (q_{++}^2(\vartheta) + q_{--}^2(\vartheta) + q_{-+}^2(\vartheta) + q_{+-}^2(\vartheta)) = 1. \quad (5.228)$$

Korzystając z (5.227) i (5.228) mamy:

$$B^2 \sin^2(\vartheta/2) + C^2 \cos^2(\vartheta/2) = (B^2 - C^2) \sin^2(\vartheta/2) + C^2 = 1. \quad (5.229)$$

Końcowa postać amplitud: Porównując współczynniki stojące przy odpowiednich funkcjach zmiennej ϑ po lewej i prawej stronie drugiej równości powyższego wyrażenia, otrzymujemy:

$$B^2 = C^2 = 1, \quad (5.230)$$

co po wstawieniu do (5.227) daje ostatecznie rozwiązanie równania generującego (5.193):

$$q_{++}(\vartheta) = q_{--}(\vartheta) = \pm \sin(\vartheta/2), \quad q_{-+}(\vartheta) = q_{+-}(\vartheta) = \pm \cos(\vartheta/2). \quad (5.231)$$

Wynik na prawdopodobieństwo w eksperymencie EPR-Bohm'a: Podstawienie amplitudy (5.231) do $P(S_{ab} | \vartheta) = \frac{1}{2} q_{ab}^2(\vartheta)$, (5.208), daje wynik na łączne prawdopodobieństwo otrzymania określonej kombinacji rzutów spinów przy zadanej wartości kąta ϑ :

$$P(++ | \vartheta) = P(-- | \vartheta) = \frac{1}{2} \sin^2(\vartheta/2), \quad P(+- | \vartheta) = P(-+ | \vartheta) = \frac{1}{2} \cos^2(\vartheta/2), \quad (5.232)$$

który jest przewidywaniem mechaniki kwantowej [56].

Na koniec, podstawiając wartości otrzymanych współczynników B'_{ab} oraz C'_{ab} do (5.216), otrzymujemy wartość stałej $g_{\vartheta\vartheta}$ dla jedynej składowej metryki Rao-Fishera w (5.219):

$$g_{\vartheta\vartheta}(\vartheta) = g_{\vartheta\vartheta} = 1, \quad (5.233)$$

która zgodnie (2.89) jest metryką samo-dualnego (Euklidesowego) układu współrzędnych ϑ na jedno-wymiarowej przestrzeni statystycznej \mathcal{S} .

Różnica pomiędzy przedstawionym wyprowadzeniem a analizą w [8]: Wynik (5.232) w ramach metody EFI, został oryginalnie wyprowadzony w [8], jednakże powyższy sposób wyprowadzenia różni się w dwóch miejscach. Po pierwsze, warunki brzegowe zostały ujęte w sposób bardziej przejrzysty [16], a po drugie w [8] odwołano się do własności ortogonalności wprowadzonych tam kwantowych amplitud, o czym wspomnimy na końcu rozdziału, a czego w powyższym wyprowadzeniu uniknięto, wprowadzając w to miejsce *warunek niezależności metryki Rao-Fishera od wartości parametru ϑ* .

Wnioski: W (5.171) wyznaczono łączne prawdopodobieństwo $P(S_{ab}) = \frac{1}{4}$. Z drugiej strony w (5.165) otrzymano $P(S_b) = 1/2$ (i analogicznie $P(S_a) = 1/2$), z czego wynika, że:

$$P(S_{ab}) = P(S_a) P(S_b) , \quad (5.234)$$

który to warunek oznacza niezależność spinowych zmiennych S_a oraz S_b .

Natomiast z (5.232) widać, że efekt korelacji spinu zmienia się bardzo mocno wraz z wartością kąta ϑ i w samej rzeczy porównanie (5.232) z (5.163) i (5.164) daje warunek:

$$P(S_{ab}|\vartheta) \neq P(S_a|\vartheta) P(S_b|\vartheta) . \quad (5.235)$$

Relacje (5.234) oraz (5.235) nie są jednak w sprzeczności. Istotnie, ponieważ prawdopodobieństwa $P(S_{ab})$, $P(S_a)$ oraz $P(S_b)$ są wyznaczone na skutek uśrednień po wszystkich wartościach kąta ϑ , zatem można było oczekiwać, że po dokonaniu tych uśrednień korelacja zmniejszy się. Zauważmy, że w (5.234) nie ma zależności od ϑ , co wynika z tego, że prawdopodobieństwo $P(S_{ab})$ z definicji określa równoczesne pojawienie się kombinacji rzutów spinów S_{ab} niezależnie od informacji o zmiennej kątowej ϑ . Podobnie, $P(S_a)$ oraz $P(S_b)$ określają prawdopodobieństwa odpowiadających im zdarzeń w sytuacji pozbycia się informacji o kącie ϑ .

Równanie (5.234) mówi więc, że w sytuacji uśrednienia po kącie ϑ , czyli wtedy gdy zmienna ta jest pod kontrolą, zmienne S_a oraz S_b rzutów spinów są niezależne, więc są również całkowicie nieskorelowane.

Natomiast wynik (5.235) dla eksperymentu EPR-Bohm'a zachodzi wtedy, gdy dokonujemy estymacji kąta ϑ metodą EFI, czyli w sytuacji oczywistego braku uśrednienia po ϑ .

Porównanie stwierdzeń o uśrednieniu: Powyższy wynik ma interesującą fizyczną interpretację. Mianowicie twierdzenie Ehrenfesta mówi, że średnie wartości kwantowych operatorów są równe ich klasycznym odpowiednikom. W jego świetle wyrażenie (5.234) mówi, że w przypadku uśrednienia po kącie ϑ stany splątane EPR-Bohm'a, na powrót ulegają klasycznej separacji. W analizie statystycznej mówimy, że po wyeliminowaniu wpływu zmiennej trzeciej, którą jest kąt ϑ , okazało się, że zmienne rzutów spinów S_a i S_b są nieskorelowane.

Uwaga o wyprowadzeniu Friedena [8]: Warunek niezależności wyprowadzenia formuł (5.227) od mechaniki kwantowej można by nieco osłabić, tzn. na tyle, aby statystyczność teorii była dalej widoczna. Frieden uczynił to, jak następuje [8]:

Pokażmy, że $q_{ab}(\vartheta)$ jest proporcjonalna do "kwantowej" amplitudy $\psi_{ab}(\vartheta)$, parametryzowanej parametrem ϑ dla stanów (5.152). Niech $P(S_{ab}, \vartheta)$ jest prawdopodobieństwem pojawienia się konfigu-

racji (a, b) rzutu spinów, *podczas gdy* parametr wynosi ϑ . Nie jest to prawdopodobieństwo łączne zajścia zdarzenia: “pojawiła się konfiguracji (a, b) rzutu spinów oraz kąt ϑ ”, gdyż ten ostatni nie jest zmienną losową.

Funkcję $\psi_{ab}(\vartheta)$ określimy tak, że kwadrat jej modułu spełnia związek:

$$|\psi_{ab}(\vartheta)|^2 \equiv |\psi_{ab}(\vartheta | S_{ab})|^2 \equiv p(\vartheta | S_{ab}) = \frac{P(S_{ab}, \vartheta)}{P(S_{ab})}, \quad (5.236)$$

tzn. jest prawdopodobieństwem, że skoro pojawiłaby się łączna konfiguracja spinów S_{ab} , to wartość kąta wynosi ϑ .

Problem polega na tym, że $q_{ab}(\vartheta)$ opisuje losowe zachowanie zmiennej spinowej S_a , a nie ϑ , co oznacza, że $q_{ab}(\vartheta)$ jest amplitudą typu $\psi_{\vartheta}(ab)$, a nie $\psi_{ab}(\vartheta)$.

Aby pokazać, że $q_{ab}(\vartheta)$ jest proporcjonalna do $\psi_{ab}(\vartheta)$, skorzystajmy z (5.236), a następnie z twierdzenia Bayesa oraz definicji prawdopodobieństwa warunkowego:

$$\begin{aligned} |\psi_{ab}(\vartheta)|^2 &= \frac{P(S_{ab}, \vartheta)}{P(S_{ab})} = \frac{P(S_{ab} | \vartheta) r(\vartheta)}{P(S_{ab} | S_b) P(S_b)} = \frac{P(S_{ab} | \vartheta) r(\vartheta)}{P(S_a | S_b) P(S_b)} = \frac{P(S_a | S_b, \vartheta) P(S_b) r(\vartheta)}{P(S_a | S_b) P(S_b)} \\ &= \frac{q_{ab}^2(\vartheta) r(\vartheta)}{P(S_a | S_b)}. \end{aligned} \quad (5.237)$$

Korzystając z (5.157), (5.172) oraz (5.175), otrzymujemy więc z (5.237), że

$$|\psi_{ab}(\vartheta)|^2 = \frac{1}{\pi} q_{ab}^2(\vartheta), \quad (5.238)$$

skąd

$$\psi_{ab}(\vartheta) = \frac{e^{i\alpha}}{\pi} q_{ab}(\vartheta), \quad \alpha \in \mathbf{R}. \quad (5.239)$$

Zatem, $\psi_{ab}(\vartheta) \propto q_{ab}(\vartheta)$, co oznacza, że¹², skoro pojawiła się łączna konfiguracja spinów S_{ab} , to amplituda prawdopodobieństwa $\psi_{ab}(\vartheta)$, że wartość kąta wynosi ϑ , jest proporcjonalna do amplitudy prawdopodobieństwa $q_{ab}(\vartheta)$ zaobserwowania rzutu spinu S_a cząstki 1 pod warunkiem, że rzut spinu cząstki 2 wynosiłby S_b , a kąt ϑ . Zdanie to jest wyrazem splątania, które pojawiło się w kwantowo-mechanicznym opisie eksperymentu EPR-Bohm’a.

W koncu, ze względu na (5.238), amplitudy $q_{ab}(\vartheta)$ są proporcjonalne do amplitud $\psi_{ab}(\vartheta)$. Zatem Frieden zarządził ortogonalności amplitud kwantowych ψ_{++} i ψ_{+-} , skąd automatycznie wyniknęła ortogonalność amplitud q_{++} oraz q_{+-} :

$$\int_0^{2\pi} d\vartheta q_{++} q_{+-} = \int_0^{2\pi} d\vartheta B'_{++} \sin(\vartheta/2) \left[B'_{+-} \sin(\vartheta/2) + C'_{+-} \cos(\vartheta/2) \right] = 0, \quad (5.240)$$

co pozwoliło na wyprowadzenie zerowania się B'_{+-} oraz B'_{-+} , (5.224), a dalej (postępując już jak powyżej) otrzymanie formuł (5.232).

¹²W mechanice kwantowej powiedzielibyśmy, że $\psi_{ab}(\vartheta)$ oznacza amplitudę prawdopodobieństwa zdarzenia, że wartość kąta wynosi ϑ , o ile pojawiłaby się łączna konfiguracja spinów S_{ab} .

5.2.4 Niepewność wyznaczenia kąta

Istnieje jeszcze jedna sprawa dotycząca analizy estymacyjnej, opisanej powyższymi rachunkami, która wymaga podkreślenia.

Rozkład eksperymentalny: Otóż w rzeczywistym pomiarze badacz otrzymuje wartości rzutu spinu S_a z określonymi częstościami, które są oszacowaniami rozkładu prawdopodobieństwa (5.232), co z kolei pozwala na punktowe oszacowanie kąta ϑ .

Analiza statystyczna metody EFI, która doprowadziła do (5.232) jest, zgodnie z postulatem EFI, sprawą układu dokonującego próbkowania przestrzeni pomiarowej położeń i estymacji oczekiwanych parametrów¹³. *Jednak jako metoda statystyczna i ona podlega pod ograniczenie Rao-Cramera do dokładności oszacowania estymowanego parametru, którym w tym przypadku jest ϑ .* Sytuacja ta ma następujące konsekwencje.

Wewnętrzny błąd estymacji metody EFI parametru ϑ : Zauważmy, że nierówność Rao-Cramera ma postać, $\sigma_{\hat{\vartheta}}^2 F \geq 1/I_F(\vartheta)$, (2.115), gdzie tym razem estymator $F \equiv \hat{\vartheta}$. Z nierówności tej otrzymujemy warunek na wariancję estymatora $\hat{\vartheta}$ kąta ϑ :

$$\sigma^2(\hat{\vartheta}) \geq \frac{1}{I_F(\vartheta)}, \quad (5.241)$$

gdzie $I_F(\vartheta) = I_b$ jest informacją Fishera (5.178), a nie pojemnością informacyjną I_{1_a} , (5.180), czy w konsekwencji jej wartością minimalną (5.206). Bez analizowania postaci estymatora $\hat{\vartheta}$, nierówność Rao-Cramera (5.241) określa DORC dla jego wariancji, o ile tylko estymator ten jest nieobciążony.

Wiemy też, że ponieważ pojemność informacyjna jest sumą po kanałach z odpowiadających im informacji Fishera, więc:

$$I_{1_a} \geq I_F(\vartheta). \quad (5.242)$$

Z (5.242) oraz z (5.241) otrzymujemy:

$$\frac{1}{I_{1_a}} \leq \frac{1}{I_F(\vartheta)} \leq \sigma^2(\hat{\vartheta}). \quad (5.243)$$

Ponieważ pojemność kanału I_{1_a} według (5.206) wynosi:

$$I_{1_a} = 4\pi, \quad (5.244)$$

więc podstawiając tą wartość do nierówności (5.243) otrzymujemy:

$$\sigma^2(\hat{\vartheta}) \geq \frac{1}{4\pi} \approx 0.08 \text{ rad}^2. \quad (5.245)$$

Wniosek: Nierówność (5.245) stwierdza, że obserwacja jednej wartości rzutu spinu przy *kompletnej nieznanowości kąta ϑ* (stąd (5.157)) daje o nim małą, lecz skończoną informację. Błąd jego estymacji $\sqrt{0,08} \text{ rad} = 0,28 \text{ rad}$ jest dość duży, co jest związane z płaską funkcją “niewiedzy” $r(\vartheta)$ określoną w (5.157).

¹³Jest tak w całej analizie EFI nie uwzględniającej wpływu urządzenia pomiarowego.

5.2.4.1 Wpływ zaszumienia pomiaru

Analiza EFI w obliczu pomiaru i płynącego z tego faktu zaszumienia danych wewnętrznych EFI nie jest przedmiotem tego skryptu. Zainteresowany czytelnik znajdzie omówienie tego tematu w [8, 16]. Poniżej zamieszczam wniosek z analizy Mroziakiewicz [16] dotyczący tego problemu dla powyższej analizy eksperymentu EPR-Bohm’a.

W pomiarze stanu układu przez zewnętrznego obserwatora, otrzymujemy dane zaszumione przez układ pomiarowy. Tzn. dane pomiarowe rzutu spinu (oznaczymy je \bar{S}_a) są generowane przez prawdziwe wartości wielkości S_a, S_b, ϑ w obecności szumu aparatury pomiarowej. Szum ten powstaje co prawda w urządzeniach Sterna-Gerlacha a oraz b , ale ze względu na założenie podane na samym początku Rozdziału 5.2.1, z innych przyczyn niż (przyjęte jako równe zero) fluktuacje rzutu spinu.

IF uwzględniająca zaszumienie: Uzyskana przy tych danych informacja Fishera I_{zasz} o parametrze ϑ , uwzględniająca zaszumienie pomiarowe, jest generowana przez informację Fishera $I_F(\vartheta) = I_b$ (5.178). W tym sensie *informacja Fishera $I_F(\vartheta)$, (5.178), procedury estymacyjnej EFI jest częścią informacji, która przejawia się w pomiarze*. Zatem obok nierówności (5.242) zachodzi również:

$$I_F(\vartheta) \geq I_{zasz}(\vartheta) . \quad (5.246)$$

Ze względu na to, że $I_{zasz}(\vartheta)$ wchodzi w (5.243) w miejsce $I_F(\vartheta)$, warunek ten oznacza pogorszenie jakości estymacji w porównaniu z (5.245).

5.2.5 Informacja Q jako miara splątania

Tak jak we wszystkich problemach estymacyjnych rozwiązanych metodą EFI, tak i w eksperymencie EPR-Bohm’a wykorzystano przy estymacji kąta ϑ , obserwowaną zmodyfikowaną informacyjną zasadę strukturalną $\tilde{i}' + \tilde{\mathbf{C}} + \kappa \mathbf{q} = 0$, (3.43), która ze względu na $\kappa = 1$ w eksperymencie EPR-Bohm’a¹⁴, daje na poziomie oczekiwanym związek $Q_{1_a} = -I_{1_a}$, (5.203), zgodnie z ogólnym warunkiem (3.45) metody EFI.

Dostępność pomiarowa I_{1_a} : Ponieważ nie mierzony był stan cząstki 2, tzn. rzut spinu S_b w czasie naszej analizy mógł przyjąć wartość $+$ lub $-$, zatem pojemność informacyjna I_{1_a} jest informacją o kącie ϑ , zawartą w obserwacji rzutu spinu cząstki 1 dokonanej przez układ i zgodnej z dokładnością do szumu z obserwacją dokonaną przez zewnętrznego obserwatora¹⁵.

Nieobserwowalna wewnętrzna struktura układu odbita w Q_{1_a} : Natomiast informacja strukturalna Q_{1_a} (5.181) dla cząstki 1, jako pozostała część informacji fizycznej K , jest informacją zawartą w nieobserwowanej wewnętrznej strukturze informacyjnej całego układu. Sugeruje to istnienie łącznej, *nieseparowalnej* informacji strukturalnej dla cząstek 1 i 2. Zatem informacja o łącznej strukturze układu, to nie to samo co suma informacji o jego składowych.

Splątanie stanów cząstek: Ze względu na $Q_{1_a} = -I_{1_a}$, przestrzeń danych odbita w I_{1_a} , tzn. obserwacje częstości rzutu spinu cząstki 1, jest na poziomie informacji, związana z wewnętrzną przestrzenią

¹⁴Dodatkowo dla eksperymentu EPR-Bohm’a $\tilde{\mathbf{C}} \equiv \tilde{\mathbf{C}}_{EPR} = 0$, zgodnie z (5.198).

¹⁵Zaszumienie z aparatury Sterna-Gerlacha, o którym wspomniano powyżej, dla uproszczenia pomijamy.

konfiguracyjną układu odbitą w Q_{1_a} . Co prawda, ze względu na brak bezpośredniego wglądu w relacje wewnątrz układu, możemy jedynie ze skończoną dokładnością (Rozdział 5.2.4) wnioskować o kącie ϑ , jednak sam fakt możliwości takiego wnioskowania określa sytuację nazywaną *splątaniem stanów* obu cząstek.

Wniosek ogólny: Wynik ten prowadzi do stwierdzenia, że strukturalna (wewnętrzna) zasada informacyjna $I = -\kappa Q$, stosowana dla każdego rozważanego problemu EFI, opisuje¹⁶ splątanie przestrzeni danych obserwowanych¹⁷ z nieobserwowaną konfiguracją układu cząstek¹⁸. Postuluje to wykorzystywanie EFI jako formalizmu do predykcji wystąpienia stanów splątanych i to nie tylko w przypadku problemu EPR-Bohm’a. Tak więc, informacja strukturala Q jest informacją o „splątaniu” widocznym w korelacji danych przestrzeni pomiarowej z przestrzenią konfiguracyjną układu.

Na koniec wspomnijmy, że inne traktowanie informacji fizycznej K w [8] niż w obecnym skrypcie, sprawia, że wnioski tam otrzymane są inne¹⁹.

¹⁶W przypadku braku separowalności I oraz Q na sumy odpowiednio pojemności informacyjnych oraz informacji strukturalnych, właściwych dla podukładów.

¹⁷Na przykład, częstości rejestracji określonych rzutów spinu S_a w eksperymencie EPR-Bohm’a.

¹⁸Na przykład, kątem ϑ w eksperymencie EPR-Bohm’a.

¹⁹Wnioski w pracy [8] są następujące: Przekaz [związanej] informacji J do *cząstki obserwowanej* wzbudza wartość spinu S_a do poziomu danej w próbce, zgodnie z końcowym prawem (5.232). Oznacza to, że przekaz informacji sprawia, że cząstka *spełnia* końcowe prawo (5.232) [opisujące zachowanie się] spinu. Przyczyną jest jakiś nieznanym mechanizm oddziaływania, być może pewna „bogata i złożona struktura, która może odpowiadać na informację i kierować zgodnie z tym, swoim własnym ruchem”. W ten sposób, przekaz [związanej] informacji J wiąże z sobą stany spinowe nieobserwowanej i obserwowanej cząstki. A zatem J (zgodnie z wymogiem) „nadaje postać” własnościom spinu *nieobserwowanej* cząstki.

Rozdział 6

Zakończenie

Celem skryptu była prezentacja uogólnienia MNW oraz zastosowania IF w jej części wnioskowania statystycznego dotyczącego estymacji niepartametrycznej metodą EFI, zaproponowaną przez Friedena i Soffera do opisu zjawisk fizycznych i ekonofizycznych. Podstawy takiej analizy statystycznego opisu zjawisk zostały wprowadzonych w latach 20 ubiegłego wieku przez Fishera.

Początkowo Fisher wprowadził MNW poszerzoną o pojęcie IF w celu rozwiązania problemu estymacji punktowej i przedziałowej parametru rozkładu zmiennej losowej w sytuacji małej próby. Jego statystyczna metoda doboru modeli, konstruowana niezależnie od ówczesnie rozwijanych teorii fizycznych, okazała się jednak sięgać o wiele dalej, w obszar estymacji równań fizycznych teorii pola, przy czym okazało się, że wielkość (małej) próby jest w tej estymacji cechą charakterystyczną modeli. Celowi temu poświęcona jest główna część niniejszego skryptu.

Jesteśmy w punkcie, w którym można już dać pewne wstępne podsumowanie metody EFI jako procedury statystycznej budowania modeli. Jest więc EFI metodą statystycznej estymacji równań ruchu teorii pola lub równań generujących rozkłady fizyki statystycznej. Gdy szukane równanie ruchu, wyestymowane metodą EFI poprzez rozwiązanie jej równań (3.43) oraz (3.44), zostało otrzymane, wtedy może być dalej prowadzone poszukiwanie rozkładu prawdopodobieństwa na analitycznych warstwach przestrzeni statystycznej \mathcal{S} , tzn. na (pod)rozmaitości w \mathcal{S} z metryką Rao-Fishera.

Może się zdarzyć, że łączny rozkład prawdopodobieństwa jest analizowany przez EFI na całej położeniowo-pędowej przestrzeni fazowej układu¹ i to dla więcej niż jednego pola. Ma to np. miejsce, gdy obok pola Diraca, obecne jest również pole cechowania (Rozdział 4.3.3.1). Wtedy na warstwie rozkładu fermionowego, w jego pochodnych kowariantnych, pole cechowania² musi być samospójnie wzięte pod uwagę.

Sedno metody EFI jest zawarte w ogólnej postaci informacji fizycznej K zadanej przez równania (4.6) oraz (4.7) (lub (4.8)). Jest ona funkcją zarówno obserwowanej informacji strukturalnej $\mathfrak{F}_n(q_n(\mathbf{x}))$ jak i amplitud $q_n(\mathbf{x})$ (lub $\psi_n(\mathbf{x})$) rozkładów, wraz z ich pochodnymi. Przy tak ogólnym zrozumieniu K , różnorodność równań EFI jest konsekwencją różnych warunków wstępnych dyktowanych przez fizykę. Mogą się one wyrażać np. poprzez równania ciągłości, które same są wynikiem estymacji statystycznej [10] (Rozdział 3.4), pewne symetrie charakterystyczne dla zjawiska (Rozdział 2.7.1) oraz warunki normalizacyjne. Wstępne założenia fizyczne mogą, poza ewentualnością nałożenia dodatkowych równań ciągłości, wskazywać również na zastosowanie jedynie wariacyjnej zasady informacyjnej, jak to

¹Gdyby na przykład chcieć wyznaczyć metodą EFI łączny rozkład energii i pędu cząstki gazu.

²Dla brzegowego rozkładu pola cechowania wyznaczonego z łącznego rozkładu wszystkich pól.

ma miejsce w przypadku równania Kleina-Gordona dla pola skalarnego (Rozdział 4.3.2), bądź na obie zasady i ich samospójne rozwiązanie, jak to miało miejsce w pozostałych przypadkach.

Podsumowując, przeszliśmy w skrypcie przez trzy etapy statystycznej rozbudowy zastosowania MNW oraz IF. Na pierwszym, wstępnym etapie, przedstawiono zastosowanie MNW oraz IF w analizie doboru modeli (Rozdział 1) oraz omówiono podstawy podejścia geometrii różniczkowej do konstrukcji przestrzeni statystycznej, obrazując poznany aparat statystyczny przykładami estymacji modeli eksponentialnych. Na drugim etapie, wprowadzono pojęcie entropii względnej i pojemności informacyjnej oraz wprowadzono strukturalną i wariacyjną zasadę informacyjną (Rozdział 2-3) metody EFI, a w trzecim przedstawiono wykorzystanie tych pojęć do wyprowadzenia podstawowych równań modeli fizycznych (Rozdziały 4-5, Dodatek 7.3.1-7.3.2), włączając w to przykład ekonofizycznego opisu zjawiska (Rozdział 5.1.4).

Rozdział 7

Dodatki

7.1 Dodatek: Zasada nieoznaczoności Heisenberga

Poniższe wyprowadzenie zasady nieoznaczoności Heisenberga zostało przedstawione w [8, 19]. Zasada Heisenberga stwierdza, że w określonej chwili czasu t nie można z dowolną dokładnością wyznaczyć jednocześnie położenia y i pędu y_p cząstki, tzn. że położenie i pęd cząstki są *rozmyte*, co możemy zapisać następująco:

$$\sigma_\theta^2 \cdot \sigma_p^2 \geq \left(\frac{\hbar}{2}\right)^2, \quad (7.1)$$

gdzie σ_θ^2 i σ_p^2 są wariancjami położenia i pędu cząstki względem ich wartości oczekiwanych θ oraz θ_p .

Wyprowadzenie nierówności (7.1) dla pola rangi $N = 2$

Założmy, że dokonujemy estymacji tylko w jednym przestrzennym kanale informacyjnym przy założeniu, że pozostałe czasoprzestrzenne parametry rozkładu są znane. Powyższa relacja może być wyprowadzona przy odwołaniu się do własności informacji Fishera. Wartości położenia y zmiennej Y są *rozmyte* (bądź fluktuują) wokół jej wartości oczekiwanej θ , skąd wartości $\mathbf{x} \in \mathcal{X}$ zmiennej odchył X spełniają (jak zwykle w skrypcie), związek:

$$\mathbf{x} = \mathbf{y} - \theta. \quad (7.2)$$

Założmy, że jest spełniona nierówność Rao-Cramera (2.115):

$$\sigma_\theta^2 \cdot I_F(\theta) \geq 1, \quad \text{gdzie} \quad \sigma_\theta^2 \equiv E \left((\hat{\theta}(Y) - \theta)^2 \right), \quad (7.3)$$

gdzie $I_F(\theta)$ jest informacją Fishera parametru θ dla *pojedynczego pomiaru* (z powodu odwołania się do skalarnej wersji twierdzenia Rao-Cramera), która po skorzystaniu z postaci kinematycznej (3.51) w Rozdziale 3.3, ma postać:

$$I_F = 4 \int_{\mathcal{X}} d\mathbf{x} \left(\frac{\partial q(\mathbf{x})}{\partial \mathbf{x}} \right)^2. \quad (7.4)$$

Rozważmy zespoloną amplitudę ψ , (3.53), dla pola rangi $N = 2$:

$$\psi(\mathbf{x}) = \psi_1(\mathbf{x}) = \frac{1}{\sqrt{2}} (q_1(\mathbf{x}) + i q_2(\mathbf{x})), \quad (7.5)$$

która ma jako część rzeczywistą i urojoną dwie rzeczywiste amplitudy q_i , $i = 1, 2$. Załóżymy, że $N = 2$ - wymiarowa próba dla zmiennej odchylen X jest prosta, skąd amplitudy $q_i(\mathbf{x})$, $i = 1, 2$, są takie same i równe $q(\mathbf{x})$:

$$q_1(\mathbf{x}) = q_2(\mathbf{x}) = q(\mathbf{x}) . \quad (7.6)$$

Zespoloną funkcję $\psi(\mathbf{x})$ można zapisać poprzez jej transformatę Fouriera, przechodząc z reprezentacji położeniowej \mathbf{x} do pędowej \mathbf{p} :

$$\psi(\mathbf{x}) = \frac{1}{\sqrt{2\pi\hbar}} \int d\mathbf{p} \phi(\mathbf{p}) \exp(i\mathbf{p}\mathbf{x}) , \quad (7.7)$$

gdzie \mathbf{p} jest rozmyciem (fluktuacją) pędu \mathbf{y}_p wokół jego wartości oczekiwanej θ_p , tzn.:

$$\mathbf{p} = \mathbf{y}_p - \theta_p , \quad (7.8)$$

a pęd cząstki jest mierzony w tej samej chwili co jego położenie.

Dla amplitudy $N = 2$ pojemność informacyjna I , (3.60), ma postać:

$$I = 8 \int d\mathbf{x} \frac{d\psi^*(\mathbf{x})}{d\mathbf{x}} \frac{d\psi(\mathbf{x})}{d\mathbf{x}} = 8 \int d\mathbf{x} \left| \frac{d\psi(\mathbf{x})}{d\mathbf{x}} \right|^2 . \quad (7.9)$$

Przedstawiając ψ w postaci:

$$\psi(\mathbf{x}) = |\psi(\mathbf{x})| \exp(iS(\mathbf{x})) \quad (7.10)$$

i wykonując różniczkowanie:

$$\frac{d\psi(\mathbf{x})}{d\mathbf{x}} = \frac{d|\psi(\mathbf{x})|}{d\mathbf{x}} e^{iS(\mathbf{x})} + i|\psi(\mathbf{x})| e^{iS(\mathbf{x})} \frac{dS(\mathbf{x})}{d\mathbf{x}} , \quad (7.11)$$

możemy (7.9) przekształcić następująco:

$$\begin{aligned} I &= 8 \int d\mathbf{x} \left| \frac{d\psi(\mathbf{x})}{d\mathbf{x}} \right|^2 = 8 \int d\mathbf{x} \frac{d\psi^*(\mathbf{x})}{d\mathbf{x}} \frac{d\psi(\mathbf{x})}{d\mathbf{x}} = \\ &= 8 \int d\mathbf{x} \left(\frac{d|\psi(\mathbf{x})|}{d\mathbf{x}} e^{-iS(\mathbf{x})} - i|\psi(\mathbf{x})| e^{-iS(\mathbf{x})} \frac{dS(\mathbf{x})}{d\mathbf{x}} \right) \left(\frac{d|\psi(\mathbf{x})|}{d\mathbf{x}} e^{iS(\mathbf{x})} + i|\psi(\mathbf{x})| e^{iS(\mathbf{x})} \frac{dS(\mathbf{x})}{d\mathbf{x}} \right) = \\ &= 8 \int d\mathbf{x} \left[\left(\frac{d|\psi(\mathbf{x})|}{d\mathbf{x}} \right)^2 + i|\psi(\mathbf{x})| e^{iS(\mathbf{x})} \frac{d|\psi(\mathbf{x})|}{d\mathbf{x}} \frac{dS(\mathbf{x})}{d\mathbf{x}} - i|\psi(\mathbf{x})| e^{iS(\mathbf{x})} \frac{dS(\mathbf{x})}{d\mathbf{x}} \frac{d|\psi(\mathbf{x})|}{d\mathbf{x}} + \right. \\ &\quad \left. + |\psi(\mathbf{x})|^2 \left(\frac{dS(\mathbf{x})}{d\mathbf{x}} \right)^2 \right] = 8 \int d\mathbf{x} \left[\left(\frac{d|\psi(\mathbf{x})|}{d\mathbf{x}} \right)^2 + |\psi(\mathbf{x})|^2 \left(\frac{dS(\mathbf{x})}{d\mathbf{x}} \right)^2 \right] . \end{aligned} \quad (7.12)$$

Zajmijmy się teraz pierwszym składnikiem pod ostatnią całką w (7.12). Ponieważ norma amplitudy ψ jest równa:

$$|\psi| = \sqrt{\frac{1}{2} (q_1^2 + q_2^2)} , \quad (7.13)$$

zatem:

$$\left(\frac{d|\psi(\mathbf{x})|}{d\mathbf{x}} \right)^2 = \frac{1}{2} \left(\frac{d}{d\mathbf{x}} \sqrt{q_1^2 + q_2^2} \right)^2 = |q_1 = q_2 = q|^2 = \left(\frac{dq}{d\mathbf{x}} \right)^2 . \quad (7.14)$$

Podstawiając powyższy wynik do całki (7.12) i korzystając z (7.4), otrzymamy następującą równość:

$$\begin{aligned} I &= 8 \int d\mathbf{x} \left| \frac{d\psi(\mathbf{x})}{d\mathbf{x}} \right|^2 = 8 \int d\mathbf{x} \left[\left(\frac{d|\psi(\mathbf{x})|}{d\mathbf{x}} \right)^2 + |\psi(\mathbf{x})|^2 \left(\frac{dS(\mathbf{x})}{d\mathbf{x}} \right)^2 \right] \\ &= 8 \int d\mathbf{x} \left(\frac{dq}{d\mathbf{x}} \right)^2 + 8 \int d\mathbf{x} |\psi(\mathbf{x})|^2 \left(\frac{dS(\mathbf{x})}{d\mathbf{x}} \right)^2 = 2I_F + 8E \left[\left(\frac{dS(\mathbf{x})}{d\mathbf{x}} \right)^2 \right] . \end{aligned} \quad (7.15)$$

Z (7.15) oraz (7.9) wynika więc następująca nierówność:

$$2I_F \leq I \quad \text{lub} \quad I_F \leq 4 \int d\mathbf{x} \left| \frac{d\psi(\mathbf{x})}{d\mathbf{x}} \right|^2. \quad (7.16)$$

Wykorzystując transformatę Fouriera (7.7), otrzymujemy z (7.16) po przejściu do reprezentacji pędowej:

$$I_F \leq \frac{4}{\hbar^2} \int d\mathbf{p} |\phi(\mathbf{p})|^2 \mathbf{p}^2. \quad (7.17)$$

gdzie $|\phi(\mathbf{p})|^2$ jest brzegową gęstością prawdopodobieństwa $P(\mathbf{p})$ odchyień pędu.

Zatem całka po prawej stronie (7.17) jest wartością oczekiwaną $E(\mathbf{p}^2)$ dla \mathbf{p}^2 :

$$I_F \leq \frac{4}{\hbar^2} E(\mathbf{p}^2) = \left(\frac{2}{\hbar} \right)^2 E[(y_p - \theta_p)^2] \equiv \left(\frac{2}{\hbar} \right)^2 \sigma_p^2, \quad (7.18)$$

gdzie σ_p^2 jest wariancją pędu cząstki, a pierwsza równość wynika z tego, że \mathbf{p} jest odchyleniem (fluktuacją) pędu y_p od jego wartości oczekiwanej θ_p , (7.8). Podstawiając powyższy wynik do nierówności Rao-Cramera, $\sigma_\theta^2 \cdot I_F(\theta) \geq 1$, (7.3), otrzymujemy (7.1):

$$\sigma_\theta^2 \cdot \sigma_p^2 \geq \left(\hbar/2 \right)^2. \quad (7.19)$$

c.n.d.

7.2 Dodatek: Równanie Schrödingera

Równania Schrödingera wyprowadzimy jako nierelatywistyczną granicę równania Kleina-Gordona. Wyprowadzenie go jako nierelatywistycznej granicy równania Diraca byłoby merytorycznie bardziej uzasadnione [29], jednak celem poniższego wyprowadzenia jest zwrócenie uwagi na relatywistyczne pochodzenie spinu elektronu.

Rozpocniemy od separacji zmiennych czasowych i przestrzennych występujących w równaniu (4.56) dla pola rangi $N = 2$, zapisując po lewej stronie wszystkie wyrazy zawierające pochodną czasową, a po prawej pochodną po współrzędnych przestrzennych:

$$\begin{aligned} & \left(\hbar^2 \frac{\partial^2}{\partial t^2} + 2 i \hbar e \phi \frac{\partial}{\partial t} + i \hbar e \frac{\partial \phi}{\partial t} - e^2 \phi^2 \right) \psi \\ & = \left(c^2 \hbar^2 \vec{\nabla}^2 - 2 i e c \hbar \left(\vec{A} \cdot \vec{\nabla} \right) - i e c \hbar \left(\vec{\nabla} \cdot \vec{A} \right) - e^2 \vec{A}^2 - m^2 c^4 \right) \psi, \end{aligned} \quad (7.20)$$

gdzie skorzystano z wyrażenia

$$\vec{\nabla} \cdot (\vec{A} \psi) = \psi \vec{\nabla} \cdot \vec{A} + \vec{A} \cdot \vec{\nabla} \psi. \quad (7.21)$$

Następnie skorzystajmy z nierelatywistycznej reprezentacji funkcji falowej [29]:

$$\psi(\mathbf{x}, t) = \tilde{\psi}(\mathbf{x}, t) e^{-imc^2 t/\hbar}, \quad (7.22)$$

gdzie wydzieliliśmy z ψ wyraz zawierający energię spoczynkową mc^2 , otrzymując nową funkcję falową $\tilde{\psi}$. Kolejnym krokiem jest zastosowanie (7.22) w (7.20). Różniczkując (7.22) po czasie otrzymujemy:

$$\frac{\partial}{\partial t} \psi = \left(\frac{\partial}{\partial t} \tilde{\psi} - \frac{imc^2}{\hbar} \tilde{\psi} \right) e^{-imc^2 t/\hbar}, \quad (7.23)$$

a po kolejnym różniczkowaniu:

$$\frac{\partial^2}{\partial t^2} \psi = \left(\frac{\partial^2}{\partial t^2} \tilde{\psi} - 2 \frac{imc^2}{\hbar} \frac{\partial}{\partial t} \tilde{\psi} - \frac{m^2 c^4}{\hbar^2} \tilde{\psi} \right) e^{-imc^2 t/\hbar}. \quad (7.24)$$

Stosując nierelatywistyczne ($n.r$) przybliżenie:

$$\frac{E_{n.r}}{mc^2} \ll 1, \quad (7.25)$$

gdzie $E_{n.r}$ jest określone poprzez poniższe zagadnienie własne:

$$-\hbar^2 \frac{\partial^2}{\partial t^2} \tilde{\psi} = E_{n.r}^2 \tilde{\psi}, \quad (7.26)$$

możemy pominąć pierwszy wyraz po prawej stronie (7.24).

Odwołajmy się do przybliżenia stałego, słabego potencjału, dla którego zachodzi:

$$e\phi \ll mc^2, \quad (7.27)$$

oraz

$$\frac{\partial \phi}{\partial t} = 0. \quad (7.28)$$

Następnie, wstawiając (7.23) oraz (7.24) do (7.20), i wykorzystując (7.25), (7.27) oraz (7.28), otrzymujemy po przemnożeniu przez $-1/2mc^2$ i oznaczeniu $\tilde{\psi}$ jako ψ , równanie:

$$i\hbar \frac{\partial}{\partial t} \psi - e\phi\psi = -\frac{\hbar^2}{2m} \vec{\nabla}^2 \psi + \frac{ie\hbar}{mc} \vec{A} \cdot \vec{\nabla} \psi + \frac{ie\hbar}{2mc} (\vec{\nabla} \cdot \vec{A}) \psi + \frac{e^2 \vec{A}^2}{2mc^2} \psi . \quad (7.29)$$

Jest to równanie Schrödingera dla nierelatywistycznej funkcji falowej ψ z potencjałem elektromagnetycznym \vec{A} oraz ϕ .

W przypadku zerowania się części wektorowej potencjału \vec{A} i dla niezerowego potencjału skalarnego ϕ , równanie to przyjmuje znaną postać:

$$i\hbar \frac{\partial}{\partial t} \psi = -\frac{\hbar^2}{2m} \vec{\nabla}^2 \psi + e\phi\psi . \quad (7.30)$$

Pytanie: Powiedz dlaczego powyższe wyprowadzenie równania Schrödingera z równania Kleina-Gordona, świadczy o relatywistycznym charakterze spinu elektronu?

7.3 Dodatek: Rezultaty EFI dla elektrodynamiki Maxwella oraz teorii grawitacji

7.3.1 Dodatek: Pole cechowania Maxwella

Poniżej zaprezentujemy rezultat metody EFI otrzymany w zapisie Friedena-Soffera dla wyprowadzenia równań Maxwella (M). Punktem wyjścia jest pojemność informacyjna (3.65):

$$I = \sum_{n=1}^N \int_{\mathcal{X}} d^4 \mathbf{x} \frac{1}{p_n(\mathbf{x})} \sum_{\nu=0}^3 \left(\frac{\partial p_n(\mathbf{x})}{\partial \mathbf{x}_\nu} \frac{\partial p_n(\mathbf{x})}{\partial \mathbf{x}^\nu} \right).$$

Rozważmy równania ruchu Maxwella dla pola rangi $N = 4$ z amplitudą rzeczywistą q_n , $n = 1, 2, 3, 4$. Zakładamy, że pola cechowania są proporcjonalne do tych rzeczywistych amplitud [8]:

$$q_\nu(\mathbf{x}) = a A_\nu(\mathbf{x}), \quad \text{gdzie } \nu \equiv n - 1 = 0, 1, 2, 3, \quad (7.31)$$

gdzie a jest pewną stałą.

Wykorzystując metrykę Minkowskiego ($\eta^{\nu\mu}$), definiujemy amplitudy $q^\nu(\mathbf{x})$ dualne do $q_\nu(\mathbf{x})$:

$$q^\nu(\mathbf{x}) \equiv \sum_{\mu=0}^3 \eta^{\nu\mu} q_\mu(\mathbf{x}) = a \sum_{\mu=0}^3 \eta^{\nu\mu} A_\mu(\mathbf{x}) \equiv a A^\nu(\mathbf{x}), \quad \text{gdzie } \nu \equiv n - 1 = 0, 1, 2, 3, \quad (7.32)$$

gdzie wprowadzono dualne pola cechowania $A^\mu(\mathbf{x})$:

$$A^\nu(\mathbf{x}) \equiv \sum_{\mu=0}^3 \eta^{\nu\mu} A_\mu(\mathbf{x}), \quad \text{gdzie } \nu = 0, 1, 2, 3. \quad (7.33)$$

Amplitudy $q_\nu(\mathbf{x})$ są związane z punktowymi rozkładami prawdopodobieństwa $p_n(\mathbf{x})$ następująco:

$$p_n(\mathbf{x}) \equiv p_{q_\nu}(\mathbf{x}) = q_\nu(\mathbf{x}) q_\nu(\mathbf{x}) = a^2 A_\nu(\mathbf{x}) A_\nu(\mathbf{x}), \quad \text{gdzie } \nu \equiv n - 1 = 0, 1, 2, 3. \quad (7.34)$$

Widzimy więc, że w metodzie EFI dla elektrodynamiki Maxwella, indeks próby n staje się indeksem czasoprzestrzennym. Zatem postać pojemności informacyjnej musi uwzględniać fakt dodatkowej estymacji w kanałach czasoprzestrzennych, przyjmując zgodnie z ogólnymi zaleceniami Rozdziału 2.7.1, postać współmienniczą:

$$\begin{aligned} I &= \sum_{\mu=0}^3 \int_{\mathcal{X}} d^4 \mathbf{x} \eta^{\mu\mu} \frac{1}{p_{q_\mu}(\mathbf{x})} \sum_{\nu=0}^3 \left(\frac{\partial p_{q_\mu}(\mathbf{x})}{\partial \mathbf{x}_\nu} \frac{\partial p_{q_\mu}(\mathbf{x})}{\partial \mathbf{x}^\nu} \right) \\ &= 4 \sum_{\mu=0}^3 \int_{\mathcal{X}} d^4 \mathbf{x} \sum_{\nu=0}^3 \left(\frac{\partial q_\mu(\mathbf{x})}{\partial \mathbf{x}_\nu} \frac{\partial q^\mu(\mathbf{x})}{\partial \mathbf{x}^\nu} \right). \end{aligned} \quad (7.35)$$

Dalej postępujemy jak w [8]. Stosujemy obie zasady informacyjne, strukturalną (3.43) z $\kappa = 1/2$ oraz wariacyjną (3.44):

$$\tilde{i} + \tilde{\mathbf{C}} + \kappa q = 0, \quad \text{gdzie } \kappa = 1/2 \quad \text{oraz} \quad \delta_{(q)}(I + Q) = 0, \quad (7.36)$$

które rozwiązujemy samospójnie, wraz z nałożonym dodatkowo warunkiem Lorentza:

$$\partial_\mu A^\mu = 0. \quad (7.37)$$

Zgodnie z (7.35) oraz (7.32) pojemność informacyjna I jest następująca:

$$I = 4 a^2 \sum_{\mu=0}^3 \int_{\mathcal{X}} d^4 \mathbf{x} \sum_{\nu=0}^3 \left(\frac{\partial A_{\mu}(\mathbf{x})}{\partial \mathbf{x}_{\nu}} \frac{\partial A^{\mu}(\mathbf{x})}{\partial \mathbf{x}^{\nu}} \right). \quad (7.38)$$

Właściwa postać informacji strukturalnej $Q \equiv Q_M$ dla równań ruchu Maxwella, została wyprowadzona w [8].

Warunki fizyczne: W celu otrzymania zarówno Q_M , jak i wartości κ , muszą być przyjęte pewne fizyczne założenia [8]. *Po pierwsze* jest to warunek Lorentza, *po drugie*, pewna wstępna postać Q_M . *Po trzecie*, wymagane jest również założenie o braku dodatkowych źródeł pola elektromagnetycznego w przestrzeni wolnej od czterowektora prądu $j^{\mu} = (c\rho, \vec{j})$.

Drugi z tych warunków, zapisany zgodnie z notacją zawartą w (4.7), wyraża się żądaniem następującej faktoryzacji obserwowanej informacji strukturalnej:

$$a^2 A_{\nu} A^{\nu} \mathfrak{F}_{\nu}(A_{\nu}(x)) = \frac{a}{\kappa} A^{\nu} F_{\nu}(j_{\nu}), \quad (7.39)$$

która jest następnie wykorzystana w (7.36). Jeżeli funkcja $F_{\nu}(j_{\nu})$ spełnia założenie o zależności jedynie od czterowektora prądu j_{ν} [8], wtedy z zasad informacyjnych metody EFI, (7.36), wynika zarówno wartość współczynnika efektywności $\kappa = 1/2$, jak i równanie ciągłości strumienia $\partial^{\nu} j_{\nu} = 0$ [43]. W efekcie informacja strukturalna dla równania Maxwella jest równa [8]:

$$Q = Q_M \equiv -\frac{64\pi}{c} a \sum_{\mu=0}^3 \int_{\mathcal{X}} d^4 \mathbf{x} j_{\mu} A^{\mu}. \quad (7.40)$$

W końcu, dla pojemności informacyjnej I jak w (7.38) i informacji strukturalnej Q jak w (7.40) oraz dla $a = 2$, metoda EFI daje wektorowe równanie falowe w cechowaniu Lorentza:

$$\square A^{\mu} = (4\pi/c) j^{\mu}, \quad (7.41)$$

co w konsekwencji prowadzi do znanej postaci równań Maxwella dla pola magnetycznego i elektrycznego [59, 8].

Podkreślmy, że proporcjonalność $q_{\nu}(\mathbf{x}) = a A_{\nu}(\mathbf{x})$ oraz warunek normalizacji:

$$(1/4) \sum_{\nu=0}^3 \int_{\mathcal{X}} d^4 \mathbf{x} q_{\nu}^2(\mathbf{x}) = \sum_{\nu=0}^3 \int_{\mathcal{X}} d^4 \mathbf{x} A_{\nu}^2(\mathbf{x}) = 1, \quad (7.42)$$

stawiają pytanie o znaczenie lokalizacji fotonu oraz istnienie jego funkcji falowej, które są ostatnio mocno dyskutowane w literaturze dotyczącej optyki. Dyskusja zawarta w [44] wspiera głównie pogląd, który stawia równania Maxwella na tych samych podstawach co równanie Diraca w sformułowaniu pierwszej kwantyzacji. Fakt ten był wcześniej zauważony w pracy Sakurai [29]. Przyjmując interpretację funkcji falowej fotonu jako mającą te same podstawy co występująca w (4.23) funkcja falowa cząstek materialnych z masą $m = 0$, z trzeciego z powyższych *fizycznych warunków* można by zrezygnować [8].

Zauważmy, że normalizacja¹ (7.42) czteropotencjału A_ν uzgadnia wartość stałej proporcjonalności² $a = 2$ z wartością $N = 4$ dla pola świetlnego.

Normalizacja (7.42), nałożona jako warunek na rozwiązanie równania (7.41), jest spójna z narzuceniem warunku początkowego Cauchy’ego we współrzędnej czasowej (por. dyskusja w [8]). Gdy jednocześnie we współrzędnych przestrzennych narzucony jest warunek Dirichlet’a (lub Neumann’a), wtedy te mieszane czasowo-przestrzenne warunki brzegowe nie są współzmiennicze, co skutkuje tym, że rozwiązanie równania (7.41) nie jest współzmiennicze.

Jednakże tylko z mieszanymi warunkami brzegowymi rozwiązanie to jest jednoznaczne [61, 8], co dla metody EFI jest warunkiem koniecznym, gdyż jest ona metodą estymacji statystycznej. Fakt ten stoi w opozycji do przypadku, gdy warunek Dirichlet’a (lub Neumann’a) jest nałożony współzmienniczo zarówno we współrzędnych przestrzennych jak i współrzędnej czasowej, gdyż co prawda otrzymane rozwiązanie jest wtedy współzmiennicze, jednak nie jest ono jednoznaczne.

W Rozdziale 4.3.2 powiedzieliśmy, że transformacja Fouriera tworzy rodzaj samosplątania pomiędzy dwoma reprezentacjami, położeniową i pędkową, dla realizowanych wartości zmiennych układu występujących w I [8]. Strukturalna zasada informacyjna wyjaśnia to splątanie, zachodzące pomiędzy pędkowymi stopniami swobody, jako spowodowane masą układu zgodną z (4.20) lub (4.30).

Rozważmy bezmasową cząstkę, np. foton. Jeśli w zgodzie z powyższymi rozważaniami dla pola Maxwella, relacja (4.32) określająca informację Fouriera miałaby być spełniona dla cząstki o masie $m = 0$ oraz dla amplitud interpretowanych zgodnie z (7.31) i (7.42) jako charakteryzujących foton, wtedy również ze strony eksperymentu należałoby się spodziewać zarówno weryfikacji kwestii związanej z naturą funkcji falowej fotonu [44] jak i sygnaturą metryki czasoprzestrzeni. W samej rzeczy, w metryce czasoprzestrzeni Minkowskiego (2.214), zgodnie z (4.30) jedyną możliwością, aby cząstka była bezmasowa, jest zachodzenie warunku $E^2/c^2 - \vec{p}^2 = 0$ dla wszystkich jej monochromatycznych, Fourierowskich mod, o ile tylko mody te miałyby posiadać fizyczną interpretację dla cząstki bezmasowej. Warunek ten oznaczałby jednak, że mody Fourierowskie nie byłyby ze sobą splątane [62] (w przeciwieństwie do tego co zachodzi dla cząstki masowej) i w zasadzie powinna istnieć możliwość dokonania detekcji każdego indywidualnego moda rozkładu Fouriera. Zatem, jeśli częstość indywidualnego moda Fouriera impulsu świetlnego nie zostałaby zarejestrowana, to mogłoby to oznaczać, że nie jest on obiektem fizycznym. Fakt ten mógłby doprowadzić do problemów dla kwantowej interpretacji fotonu, określonego jako fizyczna realizacja konkretnego Fourierowskiego moda. Problem ten został ostatnio zauważony w związku z przeprowadzonymi eksperymentami optycznymi [44], z których wynika, że Fourierowski rozkład częstości impulsu świetlnego nie reprezentuje rzeczywistych optycznych częstości, co sugeruje, że być może rzeczywisty foton jest “ziarnem elektromagnetycznej substancji” nie posiadającym Fourierowskiego przedstawienia. [44, 40].

7.3.2 Dodatek: Metoda EFI dla teorii grawitacji

Poniżej przedstawimy jedynie główne wyniki związane z konstrukcją EFI dla teorii grawitacji. Wychodząc z ogólnej postaci pojemności (3.65) i postępując analogicznie jak powyżej dla pola Ma-

¹Normalizacja czteropotencjału A_ν zadana przez (7.42) do jedności mogłaby nie zachodzić [60]. Warunkiem koniecznym dla $q_\nu(\mathbf{x})$ jest, aby niezbędne średnie mogły być wyliczone. Porównaj tekst poniżej (3.4).

²Otrzymanej wcześniej jako wynik uzgodnienia rezultatu metody EFI z równaniami Maxwella.

xwella przy definicji amplitud dualnych, otrzymujemy amplitudową postać dla pojemności informacyjnej metody EFI. Następnie postępujemy już jak w [8], gdzie zostało podane wyprowadzenie słabej (tzn. falowej) granicy równań ruchu Einsteina, dla przypadku pól z rangą $N = 10$, a amplitudy $q_n(\mathbf{x}) \equiv q_{\nu\mu}(\mathbf{x})$ w liczbie dziesięć, są rzeczywiste i symetryczne w indeksach $\nu, \mu = 0, 1, 2, 3$. Zatem, pojemności informacyjna jest następująca:

$$I = 4 \int_{\mathcal{X}} d^4\mathbf{x} \sum_{\nu,\mu=0}^3 \sum_{\gamma=0}^3 \frac{\partial q_{\nu\mu}(\mathbf{x})}{\partial \mathbf{x}_\gamma} \frac{\partial q^{\nu\mu}(\mathbf{x})}{\partial \mathbf{x}^\gamma}, \quad (7.43)$$

gdzie amplitudy dualne mają postać $q^{\delta\tau}(\mathbf{x}) = \sum_{\nu,\mu=0}^3 \eta^{\delta\nu} \eta^{\tau\mu} q_{\nu\mu}(\mathbf{x})$. Rozwiązując samospójne równania różniczkowe obu zasad informacyjnych, strukturalnej i wariacyjnej, wraz z nałożonym warunkiem Lorentza, $\sum_{\nu=0}^3 \partial_\nu q^{\nu\mu}(\mathbf{x}) = 0$, który redukuje współczynnik efektywności do wartości $\kappa = 1/2$, otrzymujemy następującą postać informacji strukturalnej:

$$Q = -8 \int_{\mathcal{X}} d^4\mathbf{x} \frac{1}{L^4} \sum_{\nu,\mu=0}^3 \bar{h}_{\nu\mu} \left[\frac{16\pi G}{c^4} T^{\nu\mu} - 2\Lambda \eta^{\nu\mu} \right], \quad (7.44)$$

gdzie

$$\bar{h}_{\nu\mu}(\mathbf{x}) \equiv L^2 q_{\nu\mu}(\mathbf{x}), \quad (7.45)$$

natomiast $\eta_{\nu\mu}$ jest metryką Minkowskiego, G jest stałą grawitacyjną, $T_{\nu\mu}$ jest tensorem energii-pędu, Λ jest tzw. stałą kosmologiczną, a stała L jest charakterystyczną skalą, na której amplitudy $\bar{h}_{\nu\mu}$ są uśrednione.

Rozwiązanie metody EFI pojawia się w postaci równania falowego dla amplitud $\bar{h}_{\nu\mu}$:

$$\square \bar{h}_{\nu\mu} = \frac{16\pi G}{c^4} T_{\nu\mu} - 2\Lambda \eta_{\nu\mu}, \quad (7.46)$$

gdzie $\square \equiv \sum_{\nu=0}^3 \partial^\nu \partial_\nu = \sum_{\mu,\nu=0}^3 \eta^{\mu\nu} \partial_\mu \partial_\nu$ jest operatorem d'Alemberta. Równania (7.46) mają postać właściwą dla równań ruchu w granicy słabego pola, w ogólnej teorii względności. Rezultatem EFI jest również równanie ciągłości strumienia, $\sum_{\nu=0}^3 \partial^\nu T_{\nu\mu}(\mathbf{x}) = 0$, dla tensora $T_{\nu\mu}$.

Pozostaje pytanie o rozwiązanie dla dowolnie silnego pola. Jedną z odpowiedzi jest następująca. Ponieważ $\square \bar{h}_{\nu\mu}$ jest jednoznacznie liniową aproksymacją tensora rangi drugiej $2R_{\nu\mu} - g_{\mu\nu}R$, gdzie $R_{\nu\mu}$ jest tensorem Ricci'ego [63], $g_{\nu\mu}$ jest tensorem metrycznym, a $R = \sum_{\nu=0}^3 R^\nu_\nu$, zatem można by uznać, że równanie Einsteina wyłania się jako jedyne możliwe, w tym sensie, że jego linearyzacją jest równanie falowe słabego pola (7.46) otrzymane w EFI [8].

Jednakże postać rozwiązania (7.46) metody EFI mogłaby również sugerować inną selekcję przyszłego modelu grawitacji. Otóż w całym formalizmie EFI amplitudy są podstawą do definicji pola, a nie metryki czasoprzestrzeni. Zatem bardziej naturalnym wydaje się zinterpretowanie $\bar{h}_{\nu\mu}$ jako pola grawitacyjnego³. Tak więc równanie (7.46) metody EFI dla grawitacji leży bliżej innej teorii grawitacji, nazywanej “szczególną teorią względności grawitacji”, która prowadzi do efektywnej teorii grawitacji typu Logunov’a [42], co poprzez widoczny związek z teoriami cechowania czyni ją “bogatszą” niż samą ogólną teorię względności. Problem statystycznego porównania obu teorii grawitacji jest kwestią przyszłych prac.

³A nie tzw. metryki słabego pola pochodzącej od liniowej części zaburzenia metryki. Metryka słabego pola ma postać: $\bar{h}_{\nu\mu} = h_{\nu\mu} - \frac{1}{2}\eta_{\mu\nu}h$, gdzie $h = \sum_{\mu,\nu=0}^3 \eta^{\mu\nu} h_{\mu\nu}$ oraz $h_{\mu\nu} = g_{\mu\nu} - \eta_{\mu\nu}$, dla $|h_{\mu\nu}| \ll 1$.

7.4 Dodatek: Informacyjny odpowiednik drugiej zasady termodynamiki:

Twierdzenie I

Niech próba będzie $N = 1$ -wymiarowa. Podobnie jak przechodzi się z (2.186) do (2.194) w celu otrzymania dyskretnej postaci informacji Fishera⁴, tak można też pokazać, że można przejść do jej następującej dyskretnej postaci:

$$I_F = \Delta x \sum_k \frac{1}{p(x_k)} \left[\frac{p(x_k + \Delta x) - p(x_k)}{\Delta x_k} \right]^2. \quad (7.49)$$

Postępując w sposób analogiczny jak w przypadku wyprowadzenia (2.199) z (2.194), otrzymujemy:

$$I_F = -\frac{2}{(\Delta x)^2} S_H(p(x) | p(x + \Delta x)). \quad (7.50)$$

Niech zmiana rozkładu prawdopodobieństwa z $p \equiv p(x)$ w chwili t do $p_{\Delta x} \equiv p(x + \Delta x)$ następuje na skutek infinytezymalnej zmiany czasu o dt . Jeśli więc dla rozkładów p oraz $p_{\Delta x}$ entropia względna spełnia warunek:

$$\frac{dS_H(p | p_{\Delta x})(t)}{dt} \geq 0, \quad (7.51)$$

wtedy ze względu na (7.50) informacja Fishera spełnia warunek:

$$\frac{dI_F(t)}{dt} \leq 0 \quad (7.52)$$

nazywany I-twierdzeniem. Oznacza on, że informacja Fishera I_F dla parametru θ maleje monotonicznie z czasem.

Uwaga: Dowód (7.51) oraz faktu, że relacja ta istotnie prowadzi do drugiej zasady termodynamiki w rozumieniu twierdzenia H, czyli dowód, który trzeba przeprowadzić dla entropii Shannona, można znaleźć w pracach na temat dynamiki układów otwartych [64].

Wprowadzenie twierdzenia I jako mającego związek z twierdzeniem H jest jak widać nieprzypadkowe. Ale również nieprzypadkowe jest podobieństwo działające w drugą stronę. A mianowicie, pojemność informacja jest związana z rangą pola N , co było widoczne w całej treści skryptu. W istniejącej literaturze można znaleźć wyprowadzenia pokazujące, że istnieje entropijny odpowiednik zasady nieoznaczoności Heisenberga, oraz związek informacji Shannona z wymiarem reprezentacji grupy obrotów dla pól fermionowych rangi N .

⁴Pojemność informacyjna (3.65) dla $N = 1$ i skalarne parametru θ wynosi:

$$I_F = \int dx \frac{1}{p_\theta(x)} \left(\frac{\partial p_\theta(x)}{\partial x} \right)^2, \quad (7.47)$$

i jest to informacja Fishera parametru θ , gdzie informację o parametrze θ pozostawiono w indeksie rozkładu. Interesujący związek informacji Fishera z entropią Kullbacka-Leiblera pojawia się na skutek zmiany rozkładu zmiennej losowej spowodowanego nie zmianą parametru rozkładu, ale zmianą wartości x na $x + \Delta x$. Zastąpmy więc (7.47) sumą Riemanna (7.49) i wprowadźmy wielkość:

$$\delta_{\Delta x} \equiv \frac{p_\theta(x_k + \Delta x)}{p_\theta(x_k)} - 1. \quad (7.48)$$

Postępując dalej podobnie jak przy przejściu od (2.195) do (2.199) (tyle, że teraz rozkłady różnią się z powodu zmiany wartości fluktuacji x), otrzymujemy (7.50).

7.4.1 Dodatek: Temperatura Fishera

Jako kolejną ilustrację nieprzypadkowego podobieństwa rachunków informacyjnych i entropijnych rozważmy definicję temperatury Fishera dla parametru θ . Otóż, informacja Fishera parametru θ pozwala na zdefiniowanie *temperatury Fishera* T_θ związanej z estymacją tego parametru:

$$\frac{1}{T_\theta} \equiv -k_\theta \frac{\partial I_F}{\partial \theta} . \quad (7.53)$$

Tak określona temperatura T_θ jest miarą czułości informacji Fishera na zmianę parametru θ , podobnie jak temperatura Boltzmannna T (dla której zachodzi $\frac{1}{T} \equiv \frac{\partial H_B}{\partial E}$) jest miarą czułości entropii Boltzmannna H_B na zmianę energii E układu. Temperatura Fishera znalazła swoje zastosowanie między innymi w badaniu rynków finansowych [65, 66].

Bibliografia

- [1] R.A. Fisher, Phil. Trans. R. Soc. Lond. **222**, 309 (1922). R.A. Fisher, *Statistical methods and scientific inference*, 2 nd edn. (London, Oliver and Boyd, 1959).
- [2] C.R. Rao, Bulletin of the Calcutta Mathem. Soc. **37**, pp.81-91 (1945). *Information and accuracy attainable in the estimation of statistical parameters*.
- [3] B. Efron, The Annals of Statistics **3**, pp.1189-1242 (1975). *Defining the curvature of statistical problem (with application to second order efficiency) (with discussion)*.
- [4] A.P. Dawid, The Annals of Statistics **3**, pp.1231-1234 (1975). *Discussion to Efron's paper*. A.P. Dawid, The Annals of Statistics **5**, pp.1249 (1977). *Further comments on a paper by Bradley Efron*.
- [5] S. Amari, POST RAAG Report **106** (1980). *Teory of information spaces-a geometrical foundation of statistics*. S. Amari, The Annals of Statistics **10**, pp.357-385 (1982). *Differential geometry of curved exponential families-curvature and information loss*.
- [6] S. Amari, H. Nagaoka, *Methods of information geometry, translations of Mathematical monographs*, Vol.191, Oxford Univ. Press, (2000).
A. Fujiwara and H. Nagaoka, *Quantum Fisher metric and estimation for pure state models*, Phys. Lett. A **201**, 119-124, (1995). A. Fujiwara and H. Nagaoka, *An estimation theoretical characterization of coherent states*, J. Math. Phys. **40**, 4227-4239, (1999). K. Matsumoto, *A geometrical approach to quantum estimation theory*, PhD thesis, Univ. of Tokyo, (1998).
- [7] A. Fujiwara and H. Nagaoka, Phys. Lett. A **201**, pp.119-124 (1995). *Quantum Fisher metric and estimation for pure state models*. A. Fujiwara and H. Nagaoka, Journal of Mathem. Phys. **40**, pp. 4227-4239 (1999).
An estimation theoretical characterization of coherent states. K. Matsumoto, Technical Report METR 96-09, Dept. of Math. Eng. and Inform. Phys, Univ. of Tokyo, (1996). *A new approach to the Cramer-Rao type bound of the pure state model*. K. Matsumoto, PhD thesis, Univ. of Tokyo, (1998). *A geometrical approach to quantum estimation theory*.
- [8] B.R. Frieden, Found.Phys. **16**, No.9, *A probability law for the fundamental constants*, 883-903 (1986). B.R. Frieden, Phys.Rev.A **41**, *Fisher information, disorder, and the equilibrium distributions of physics*, 4265-4276 (1990). B.R. Frieden, B.H. Soffer, Phys.Rev.E **52**, *Lagrangians of physics and the game of Fisher-information transfer*, 2274-2286 (1995). B.R. Frieden, Phys.Rev.A **66**, *Relations between parameters of a decoherent system and Fisher information*, 022107, (2002). B.R. Frieden, A. Plastino, A.R. Plastino and B.H. Soffer, *Schrödinger link between nonequilibrium thermodynamics and Fisher information*, Phys.Rev.E **66**, 046128 (2002). B.R. Frieden, A. Plastino, A.R. Plastino

- and B.H. Soffer, *Non-equilibrium thermodynamics and Fisher information: An illustrative example*, *Phys.Lett.A* **304**, (2002), pp.73-78.
- B.R. Frieden, *Science from Fisher information: A unification*, (Cambridge University Press, 2004).
- [9] J. Syska, *The Fisher information and quantum models-classical field theory, classical statistics similarity*, *Phys. Stat. Sol.(b)*, **244**, No.7, 2531-2537 (2007)/DOI 10.1002/pssb.200674646.
- [10] E.W. Piotrowski, J. Śladkowski, J. Syska, S. Zając, *The method of the likelihood and the Fisher information in the construction of physical models*, *Phys. Stat. Sol.(b)*, **246**, No.5, 1033-1037, (2009)/DOI 10.1002/pssb.200881566, arXiv:physics/0811.3554.
- [11] R. Nowak, *Statystyka dla fizyków*, PWN, Warszawa 2002.
- [12] J. Jakubowski, R. Sztencel, *Wstęp do teorii prawdopodobieństwa*, wydanie 2, Script, Warszawa, (2001).
- [13] Y. Pawitan, *In all likelihood: Statistical modelling and inference using likelihood*, (Oxford Univ. Press, 2001).
- [14] D.G. Kleinbaum, L.L. Kupper, K.E. Muller, A. Nizam, *Applied Regression Analysis and Other Multivariable Method*, Duxbury Press, (1998).
- [15] D. Mroziakiewicz, *Analiza regresji Poissona z estymatorami metody największej wiarygodności z wykorzystaniem programu statystycznego SAS*, Praca licencjacka, Inst. Fizyki, Univ. Śląski (Rybnik), Polska, (2006).
- [16] D. Mroziakiewicz, *Metoda największej wiarygodności i informacja Fishera w fizyce i statystyce*, Praca mgr., (*The maximum likelihood method and Fisher information in physics and statistics*), unpublished, Uniwersytet Śląski, Inst. Fizyki, (2008);
- [17] M. Czerwik, *Wykorzystanie programu SAS jako narzędzia do analizy współzależności zmiennych metodą regresji*, Praca licencjacka, Uniwersytet Śląski, Inst. Fizyki, Jastrzębie Zdrój, (2004).
- [18] M. Biesiada, *Statystyka w ujęciu Bayesowskim*, Skrypt dla studentów ekonofizyki, Uniwersytet Śląski, Instytut Fizyki, (2011).
- J-P. Gazeau, *Coherent states in quantum physics*, Wiley-VCH, (2009).
- [19] S. Mania, *Informacja Fishera. Wyprowadzenie równania Kleina-Gordona*, Praca licencjacka, Uniwersytet Śląski, Inst. Fizyki, Jastrzębie Zdrój, (2007).
- [20] I. Bengtsson, K. Życzkowski, *Geometry of quantum states*, Cambridge Univ. Press, (2006).
- [21] R.F. Streater, *Statistical dynamics. A stochastic approach to nonequilibrium thermodynamics*, 2nd ed., Imperial College Press, (2009).
- [22] A. Ohara and T. Kitamori, *Geometric structures of stable state feedback systems*, *IEEE Transactions on Automatic Control*, **38**(10), pp.1579-1583, (1993). A. Ohara and S. Amari, *Differential geometric structures of stable state feedback systems with dual connections*, *Kybernetika*, **30**(4), pp.369-386, (1994). A. Ohara, N. Suda, and S. Amari, *Dualistic differential geometry of positive definite matrices and its applications*, *Linear Algebra and its Applications*, **247**, pp.31-53, (1996).

- [23] P.J. Brockwell, R.A. Davis, *Time Series: Theory and Methods*, 2nd edition, Springer, (2009).
Ł. Machura, *Analiza szeregów czasowych*, Skrypt dla studentów ekonofizyki, Uniwersytet Śląski, Instytut Fizyki, (2011).
- [24] M.K. Murray, J.W. Rice, *Differential geometry and statistics*, Chapman and Hall/CRC, (1993).
- [25] M. Nakahara, *Geometry, topology and physics*, 2nd edition, IOP Publishing Ltd, (2003).
- [26] A.P. Dembo and O. Zeitouni, *Large deviations techniques and applications*, 2nd ed. Springer, New York, (1998).
- [27] A.J. Stam, *Some mathematical properties of quantities of information*, Ph.D. dissertation. Technical University of Delf, (1959). A. J. Stam, *Some inequalities satisfied by the quantities of information of Fisher and Shannon*, Inform. and Control 2, 101–112 (1959).
- [28] F. Pennini, A. Plastino, B.H. Soffer, C. Vignat, *Physical symmetries and Fisher's information measure*, Phys.Lett. A373, pp.817-820, (2009).
- [29] J.J. Sakurai, *Advanced quantum mechanics*, p.169, Addison-Wesley Publishing Company, (1967).
- [30] J. Syska, *Kaluza-Klein Type Model for the Structure of the Neutral Particle-like Solutions*, International Journal of Theoretical Physics, Volume 49, Issue 9, p.2131-2157, DOI:10.1007/s10773-010-0400-8, Open Access, (2010).
- [31] J. Dajka, *Kwantowe metody opisu dynamiki instrumentów finansowych*, Skrypt dla studentów ekonofizyki, Uniwersytet Śląski, Instytut Fizyki, (2010).
- [32] I.N. Bronsztejn, K.A. Siemiendiajew, G. Musiol, H. Mühlig, *Nowoczesne kompendium matematyki*, 305-306 Wydawnictwo PWN, 305-306, (2004).
- [33] S. Zająć, *Modelowanie szeregów czasowych za pomocą procesów ARMA i ARIMA, (aplikacje w systemie SAS)*, Praca licencjacka, Inst. Fizyki, Univ. Śląski (Rybnik), Polska, (2005).
- [34] K. Sobczyk, *Stochastyczne równania różniczkowe. (Teoria i zastosowania)*, Wydawnictwo NT, (1996).
J. Łuczka, *Procesy i zjawiska losowe*, Skrypt dla studentów ekonofizyki, Uniwersytet Śląski, Instytut Fizyki, (2010).
- [35] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik*, (Springer, Berlin), 171 (1932); transl. by E.T. Beyer, *Mathematical foundations of quantum mechanics*, Princeton Univ. Press, Princeton, 324 (1955).
- [36] A. Peres, *Quantum Theory: Concepts and Methods*, Kluwer Academic Publishers, (1995).
- [37] M.R. Esfandiary, H.J. Fink, Phys.Lett.A54, 383 (1975). V.A. Schweigert, F.M. Peeters et.al., Phys.Rev.Lett.79, 4653 (1997); Phys.Rev.Lett.83, 2409 (1999); Phys.Rev.B59, 6039 (1999); Phys.Rev.B62, 9663 (2000). J.J. Palacios, Phys.Rev.B57, 10873 (1998); Phys.Rev.Lett.83, 2409 (1999); Phys.Rev.Lett.84, 1796 (2000). G.F. Zharkov, V.G. Zharkov and A.Yu. Zvetov, Phys.Rev.B61, 12293 (2000). G.F. Zharkov, Phys.Rev.B63, 214502 (2001); Phys.Rev.B63, 224513 (2001).

- [38] A.O. Barut, J.F. Van Huele, *Phys.Rev.A* **32**, 3187 (1985). A.O. Barut and N. Ünal, *J.Math.Physics* **27**, 3055 (1986). A.O. Barut, J.P. Dowling, *Phys.Rev.A* **36**, 649 (1987). A.O. Barut and N. Ünal, *Physica* **142 A**, 467, 488 (1987). A.O. Barut, *Phys.Scr.* **T21**, 18 (1988). A.O. Barut, Y.I. Salamin, *Phys.Rev.A* **37**, 2284 (1988).
J. Syska, Self-consistent classical fields in gauge field theories, PhD thesis, (University of Silesia, 1995/99), (unpublished); *Trends in Boson Research*, ed. A.V. Ling, (Nova Science Publishers), 163-181 (2006); *Frontiers in field theory*, ed. O. Kovras, (Nova Science Publishers), 125-154 (2005).
- [39] I. Bednarek, R. Mańka, *J.Phys.G* **31**, 1, (2005).
- [40] J. Syska, *Boson ground state fields in the classical counterpart of the electroweak theory with non-zero charge densities*, in *Frontiers in field theory*, ed. O. Kovras, (Nova Science Publishers), (Nowy Jork 2005), chapter 6, 125-154 and in *Focus on Boson Research*, (Nova Publishers), 233-258, (2006).
- [41] J. Syska, *Remarks on self-consistent models of a particle*, in *Trends in Boson Research*, ed. A.V. Ling, (Nova Science Publishers), 163-181, (2006).
- [42] V.I. Denisov and A.A. Logunov, *The Theory of Space-Time and Gravitation*, in *Gravitation and Elementary Particle Physics*, Physics Series, ed. A.A. Logunov, (MIR Publishers, Moscow), 14-130, (1983). A.A. Logunov, *The Theory of Gravity*, arXiv:gr-qc/0210005v2, (2002).
- [43] J. Syska, R. Szafron, *Informational criteria for the field theory models building*, in preparation.
- [44] Edited by Ch. Roychoudhuri, A.F. Kracklauer, K. Creath, *The nature of light. What is a photon?*, (CRC Press, Taylor&Francis Group), (2008).
- [45] F.del Aguila, J. Syska, R. Szafron, S. Zając, M. Zrałek, *The dynamics of the neutrino flavour transition with the general effective interaction*, in preparation.
- [46] R. Ticciati, *Quantum field theory for mathematicians*, (Cambridge University Press), (1999).
- [47] M. Fecko, *Differential geometry and Lie groups for physicists*, (Cambridge University Press), (2006).
- [48] A. Khrennikov, *Bell-Boole Inequality: Nonlocality or Probabilistic Incompatibility of Random Variables?*, *Entropy* **10**, pp.19-32, (2008).
- [49] W.I. Arnold, *Teoria równań różniczkowych*, PWN, Warszawa, (1983).
- [50] M. Aoki, H. Yoshikawa, *Reconstructing macroeconomics. A perspective from statistical physics and combinatorial stochastic processes*, (Cambridge University Press), (2007).
- [51] U. Garibaldi, E. Scalas, *Finitary Probabilistic Methods in Econophysics*, Cambridge Univ. Press, 1st ed., (2010).
- [52] J. Flachsmeier, *Kombinatoryka*, PWN, (1977).
- [53] M. Abramowitz, I. A. Stegun. *Handbook of Mathematical Functions* Washington, DC.: National Bureau of Standards, 1965.

- [54] E. Kamke, *Sprawocznik po obykowniennym differencjalnym urawnieniam*, red. C.W. Fomin, Izdatielstwo inostrannoj litieratury, Moskwa, 1951.
- [55] I.B. Rumer and M.Sh. Ryvkin, *Thermodynamics, Statistical Physics and Kinetics*, Moscow: Mir Publishers, p.526, (1980).
- [56] E.B. Manoukian, *Quantum Theory A Wide Spectrum*, Springer, (2006).
- [57] J.S. Bell, *Speakable and unspeakable in quantum mechanics*, Cambridge Univ. Press: Cambridge, UK, (1987).
- [58] L. Accardi, *The probabilistic roots of the quantum mechanical paradoxes*, Proc. Conf. The wave–particle dualism. A tribute to Louis de Broglie on his 90th Birthday, Reidel Publ. Company: Dordrecht, Netherlands, pp. 47–55, (1970). A. Fine, *Hidden Variables, Joint Probability, and the Bell Inequalities*, Phys. Rev. Lett. **48**, pp.291–295, (1982). I. Pitowsky, *Resolution of the Einstein-Podolsky-Rosen and Bell Paradoxes*, Phys. Rev. Lett. **48**, pp.1299-1302, 1982. I. Pitowsky, *Deterministic Model of Spin and Statistics*, Phys. Rev. D **27**, pp.2316-2326, (1983). P. Rastall, *The Bell inequalities*, Found. Phys. **13**, 555-575, (1983). K. Hess, W. Philipp, *A possible loophole in the theorem of Bell*, Proc. Natl. Acad. Sci. **98**, 14224, (2001). K. Hess, W. Philipp, *Exclusion of time in the theorem of Bell*, Europhys. Lett. **57**, 775, (2002). A. Khrennikov, *Statistical measure of ensemble nonreproducibility and correction to Bell's inequality*, Il Nuovo Cimento B, **115**, pp.179–184, (1999). A. Khrennikov, *Non-Kolmogorov probability models and modified Bell's inequality*, J. Math. Phys. **41**, pp.1768–1777, (2000). A. Khrennikov, *A perturbation of CHSH inequality induced by fluctuations of ensemble distributions*, J. Math. Phys. **41**, pp.5934–5944, (2000).
- [59] J. D. Jackson, *Elektrodynamika klasyczna*, PWN, Warszawa, (1982).
- [60] U. Leonhardt, *Essential quantum optics. From quantum measurements to black holes*, (Cambridge Univ.Press.), (2010).
- [61] P.M. Morse and H. Feshbach, *Methods of Theoretical Physics*, (New York: McGraw-Hill), (1953).
- [62] J. Śładkowski, J. Syska, *The information channel capacity in the field theory estimation*, sent to Phys.Lett.A, (2012).
- [63] C.W. Misner, K.S. Thorne, J.A. Wheeler, *Gravitation*, New York: Freeman, (1973).
- [64] A. Wehrl, Rev. Mod. Phys. **50**, 221, (1978); R. Alicki and K. Lendi, *Quantum Dynamical Semigroups and Applications*, Lect. Notes Phys. **717**, (Springer, Berlin Heidelberg), (2007), DOI 10.1007/b11976790.
- [65] E.W. Piotrowski, J. Śładkowski, *The thermodynamics of portfolios*, Acta Physica Polonica B, Vol.32, No.2, (2001); H. Kleinert, X.J. Chen, *Boltzmann distribution and market temperature*, arXiv:physics/0609209v2, (2007).
- [66] René Brand, *An econophysical investigation: Using the Boltzmann-distribution to determine market temperature as applied to the JSE All Share Index*, Stellenbosch University, (16 March 2009).